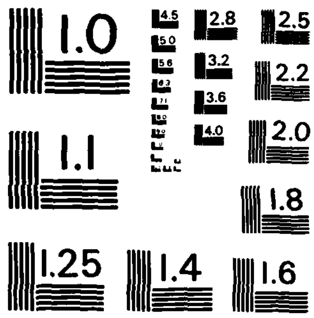
AD-A147 661 RECREATIONAL BOATING SAFETY: ANALYSIS FOR PROGRAMMATIC DECISIONS(U) MANDEX INC MCLEAN VA L GREENBERG ET AL. APR 84 USCG-D-9-84 DTCG23-83-C-20080 1/2

UNCLASSIFIED F/G 13/12 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS - 1963 - A

Report No. CG-D-9-84

RECREATIONAL BOATING SAFETY

ANALYSIS FOR PROGRAMMATIC DECISIONS

Mandex, Inc.
8302D Old Courthouse Road
Vienna, Virginia 22180

and

Decision Science Consortium, Inc. (subcontractor)
7700 Leesburg Pike, Suite 421
Falls Church, Virginia 22043

April 1984

Document is available to the public through the
National Technical Information Service,
Springfield, Virginia 22161

Prepared for

DEPARTMENT OF TRANSPORTATION
UNITED STATES COAST GUARD
Office of Research and Development
Washington, D.C. 20593

NOTICE

Technical Report Documentation Page

1. Report No. CG-D-9-84	2. Government Accession No. AD-A147 661	3. Recipient's Catalog No.
4. Title and Subtitle Recreational Boating Safety: Analysis for Programmatic Decisions		5. Report Date April 1984
		6. Performing Organization Code
7. Author(s) Leonard Greenberg, Terry A. Bresnick, Jacob W. Ulvila		8. Performing Organization Report No.
9. Performing Organization Name and Address Mandex, Inc. 8302D Old Courthouse Rd. & Vienna, VA 22180 / Decision Science Consortium 7700 Leesburg Pike #421 Falls Church, VA 22043		10. Work Unit No. (TRAIS)
		11. Contract or Grant No. DTCG23-83-C-20080
12. Sponsoring Agency Name and Address Department of Transportation, U.S. Coast Guard Office of Research and Development Washington, D. C. 20593		13. Type of Report and Period Covered Final Report Sept. 1983 - April 1984
		14. Sponsoring Agency Code G-FCP-22C/64

15. Supplementary Notes

16. Abstract

This is a users' manual for addressing analytic problems in the realm of recreational boating safety. Four major areas of application are identified: Needs Assessment, Performance Prediction, Performance Evaluation, and Resource Allocation. Analytic techniques useful in each area are identified and explained. Techniques presented include: cause assessment models and tests for non-randomness ("Needs Assessment"), data smoothing and Box-Jenkins time series analysis ("Performance Prediction"), Box-Tiao and other forms of intervention analysis ("Performance Evaluation"), and methods for assessing program cost, benefit, and utility ("Resource Allocation"). An additional technique, Multiattribute Utility (MAU) Analysis, is also described, for use in addressing situations involving multiple objectives. [illegible]

17. Key Words Recreational Boating, Regulatory Analysis, Benefit Analysis, Impact Assessment, Needs Assessment, Forecasting Models, Times Series Analysis, Program Evaluation, Resource Allocation, Multiattribute Utility Analysis	18. Distribution Statement Document available to the U.S. public through the National Technical Information Service, Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	129	

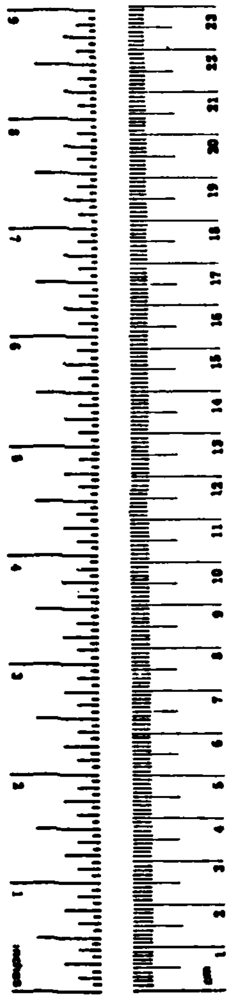
Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F

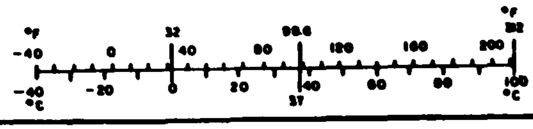


TABLE OF CONTENTS

FOREWORD

The purpose of this Users' Guide is to increase the awareness of Recreational Boating Safety (RBS) personnel concerning the analytic techniques available to support their mission, and to facilitate, through clarification and example, the use of those techniques. The Guide is designed to be read and understood by all levels of personnel -- analysts, managers, and decision-makers.

This Guide is the joint product of Mandex, Inc., and its subcontractor, Decision Science Consortium (DSC). Chapter 1 (How to Use Analysts: A Guide for Managers) was written by Jacob Ulvila of DSC. Mr. Ulvila also collaborated with Terry Bresnick of DSC in writing Chapters 3 and 4. Mr. Bresnick was the principal author of Chapter 3 on multiattribute utility analysis and co-author of Chapter 4 on resource allocation methods; Mr. Ulvila was the principal author of the latter chapter and co-author of the former. The undersigned wrote the Introduction and Chapter 2.

This effort was guided by Dr. John Gardenier, who served as Government Project Officer. Other Coast Guard personnel who shared in the development of the Guide include Commander Wayne Becker, Gary Traub, Ladd Hakes, and Lyle Gray, all of whom contributed useful insights and critiques of earlier drafts.

Typing of the Guide was performed by Frances Inman of Mandex and Diane Laaksonen of DSC. Their diligence, competence, and patience in support of this effort are deeply appreciated.

Although couched within the context of recreational boating safety, the techniques described in this Guide are, in many instances, of considerably broader applicability. It is hoped that they will be used in that broader context as well.

Leonard Greenberg
Mandex, Inc.
Project Director

INTRODUCTION

The Recreational Boating Safety (RBS) Program of the U. S. Coast Guard has as its purpose:

> "... to minimize the loss of life, personal injury, and property damage associated with the use of recreational boats, through preventive means, in order to provide the public with maximum safe use of the nation's waters."[1]

To accomplish its purpose, the RBS Program has set goals and priorities in three broad areas of program activity, designed to save lives, reduce personal injuries and minimize property damage. As set forth in the RBS Operating Program Plan, those areas of activity are as follows:

a. Boat and Equipment Safety. - To establish safety standards and guidelines for boats and associated equipment, and to ensure that those standards are met.

b. Boater Education. - To adopt programs and coordinate activities designed to improve boaters' knowledge, abilities, and attitudes.

c. Boating Environment. - To promote law enforcement and other activities generally conducive to improved boating safety.

Activities such as these present the Program with not only a charter but also an obligation. Those responsible for program policy and implementation must see to it that both the regulations and the associated educational and enforcement activities meet valid safety needs, and that these efforts yield benefits at least commensurate with their cost. Choices among alternative safety actions must be justified. Resources must be allocated in a prudent and reasonable manner, with proper attention to matters of priority, cost, and utility.

---

[1] U. S. Coast Guard, Recreational Boating Safety Operating Program Plan, FY 86-90, p. I-1.

All of this requires analysis. It requires the use of data which are both relevant to the decision-making process and (within reason) accurate, timely, and complete. Above all, it requires the use of analytic techniques designed to extract the maximum intelligence from the data and to make effective use of subjective judgment where the raw numbers are deficient or otherwise inadequate.

Boating safety analytic functions center about four major areas of application. Although these areas are interrelated, often using the same techniques and data, each addresses a fundamentally different set of concerns. These four basic areas of application, and the issues typically addressed by each, are identified below:

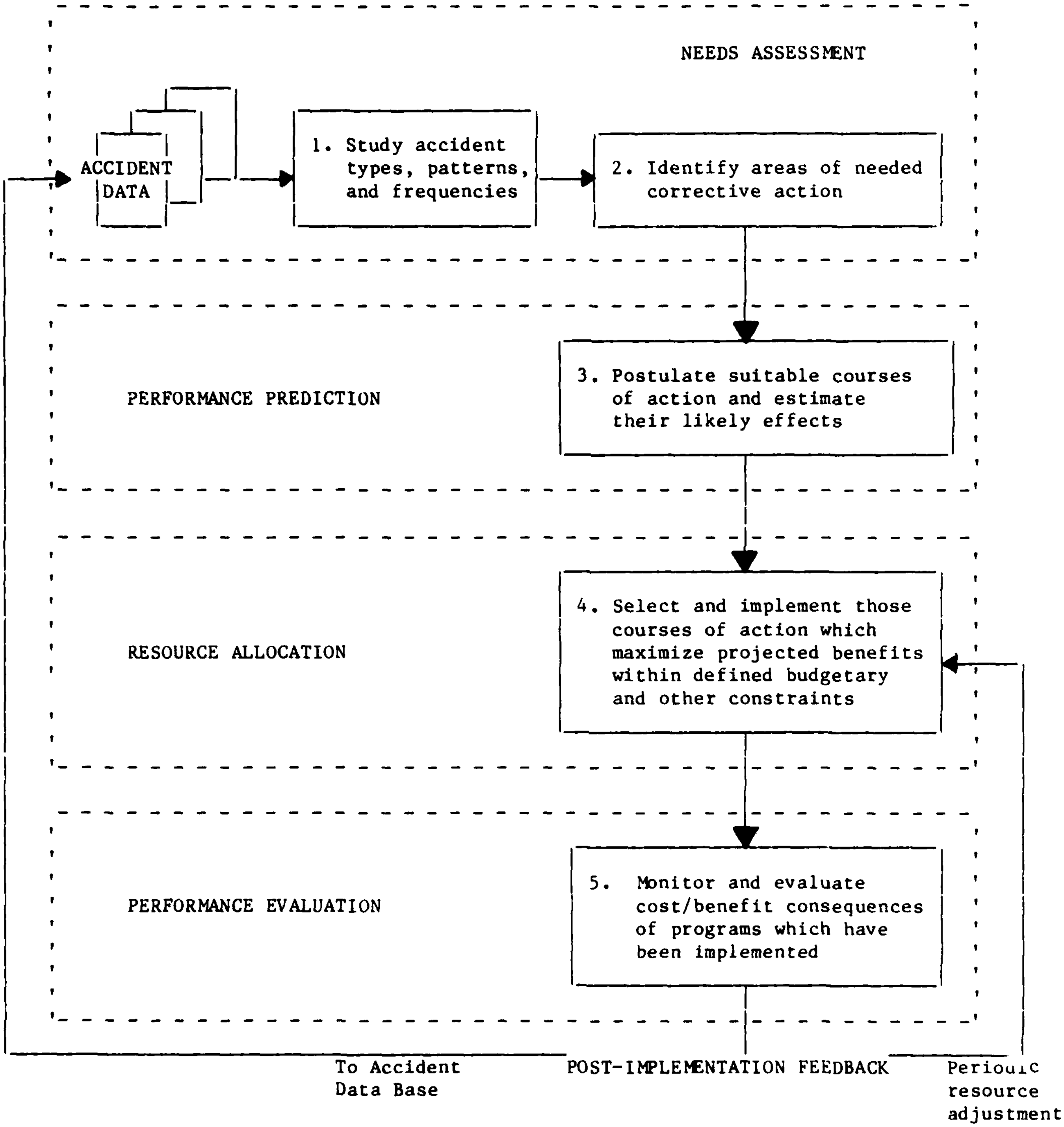
Application Area	Issues Typically Addressed
Needs Assessment	Which aspects of boating safety seem most in need of corrective action? Are there identifiable areas of boating activity where appropriate countermeasures (regulatory, educational, or otherwise) might reasonably be taken to reduce the frequency of accidents or mitigate their severity?
Performance Prediction	For each countermeasure being considered for adoption, what are its probable benefits in terms of reduced fatalities, personal injury, and/or property damage? What is its probable cost? Given two or more alternative options, which alternative is likely to produce the greatest payoff?
Performance Evaluation	Given a specific safety program currently or previously in operation, to what extent has that program achieved its stated objectives? Has the program yielded benefits commensurate with its cost? Which aspects of the program seem to have worked better than others? Which have worked poorly? What

Application Area (cont'd)	Issues Typically Addressed (cont'd)
	factors seem to be responsible for governing program failure or success?
Resource Allocation	Given finite resources and a fixed budget, which specific mix of activities is likely to produce the greatest overall utility, i.e., will best meet the defined program objectives?

The way in which these activities interrelate is shown in Figure 1. The analytic process typically starts with the study of accident data, either self-initiated or in response to external mandates and requests. As a result of this study, areas of needed corrective action are identified ("needs assessment"). Alternative measures for countering those needs are then postulated and their likely effects studied ("performance prediction"). From among these alternative possibilities, those specific courses of action which fit within budget and are deemed likely to produce optimum results are adopted and implemented ("resource allocation"). The outcome of these programs, in terms of both benefit and cost, is then systematically monitored and evaluated ("performance evaluation"). Finally, the process of resource allocation is repeated as post-implementation decisions are reached as to which programs should be continued, which should be modified, and which should be discontinued or curtailed. Feedback, based on program performance, is also provided to the accident data base.

Support of this process -- or portions of it at a time -- constitutes the analyst's primary mission. These activities necessarily entail a fair amount of quantitative analysis. The form of analysis, however, can vary considerably: techniques suitable for one purpose may not be suitable for another; techniques suitable where there are ample reliable data may not be suitable where the data are suspect or sparse; techniques suitable where there are adequate computer resources and ample time for reaching a decision may not be suitable where these resources are lacking or if quick answers are required; and so on.

Figure 1. Relationship of Analytic Processes Within the RBS Program Framework



Because of this need for (and availability of) multiple analytic methods, the Coast Guard's Office of Boating Safety initiated a multiphase R & D effort designed to describe, categorize, and illustrate the more useful of those methods. This effort, conducted by Wyle Laboratories, resulted in a draft document entitled "Recreational Boating Program Effectiveness Methodology" (hereinafter termed "the Wyle report").[1]

Because of its massive volume and highly mathematical orientation, the Wyle report has been found somewhat limited in its ability to serve the needs of a small program analysis staff with severe time constraints and less than Ph.D. level mathematical background. The present publication is therefore designed to (a) clarify and (b) supplement the Wyle report. The clarification (Chapter 2) consists of a series of discussions which deal with selected portions of the Wyle report within a generally application-oriented framework. The supplementation (Chapters 3 and 4) consists of an extended treatment of two topics relevant to the RBS mission which were not explicitly treated in the Wyle report. These additional topics are:

1. Multiattribute Utility Analysis (Chapter 3)

   Multiattribute utility (MAU) analysis is a technique useful in situations where there are multiple objectives to be met and no single solution ranks best with respect to all objectives. The essence of MAU models is twofold: first (a) each objective is assigned a weight which reflects the importance of that objective in the overall decision process, then (b) each proposed solution is assigned a series of scores which reflect how well that solution does with respect to each of the objectives. The latter set of scores is then combined with the initial set of objective weights to produce a single measure reflective of the overall utility of that solution. The solution which ranks highest in utility is normally taken, at least initially, to be

---

[1] The 816-page document produced by Wyle Laboratories was prepared in draft form only and is unlikely to be formally published, although copies are available to boating safety analysts at Coast Guard Headquarters. Completed in November 1981, the document is colloquially referred to by boating safety analysts as "the Wyle report." (Actually, Wyle Laboratories produced many other reports for the Coast Guard's RBS program, as well as for other customers.) For convenient reference by the primary intended reading audience, the shorthand term, "the Wyle report," is used herein.

the "preferred" solution. The process does lend itself, however, to additional iterations as the user attempts to refine the initial selection.

MAU methods are quite versatile. In terms of specific areas of application, MAU models work well in at least three of the four contexts defined earlier, i.e., they can be used to predict the utility of program options which have yet to be tried ("performance prediction"), to assess the utility of options which <u>have</u> been tried ("performance evaluation"), and to provide measures of utility which can be used in reaching decisions on the investment of resources ("resource allocation"). The sole requirement for the use of MAU methods is that the problem be multidimensional.

2. <u>Resource Allocation Methods</u> (Chapter 4)

These are methods which, as their name implies, are useful in connection with the fourth area of application defined earlier: Resource Allocation. The major thrust of these methods is to allocate a constrained <u>resource</u> to different competing <u>uses</u> so as to produce maximum <u>benefit</u>. The constrained resource might be money or personnel; the different uses might be alternate program activities. Benefit might be expressed in terms of a single quantity such as reduced fatalities or it might be multidimensional, in which case MAU analysis can be profitably brought into play.

The rationale by which resource allocation methods maximize benefit is by identifying, in decreasing order, that sequence of allocations which produces the greatest increment in benefit per unit of resource expended.

The material presented in this Users' Guide requires the ability to think logically and analytically, but does not demand a level of mathematical sophistication beyond that normally encountered in most college graduates. Mathematical notation is kept to a minimum and the use of tables, charts, and examples is emphasized.

Recognizing that analysis works best when there is effective communication between those who commission analyses ("managers") and those who perform them ("analysts"), Chapter 1 tells how to achieve such a productive relationship.

1. HOW TO USE ANALYSTS: A GUIDE FOR MANAGERS

Introduction

A good deal has been written about how a manager can ensure the best results from an analysis. (See Bibliography, end of this chapter.) Many managers and analysts are unaware of this literature, and much of the advice is overlooked in practice. This paper offers tips for managers and analysts gleaned both from consulting experience and from the literature.

This is a brief overview; someone desiring more extensive reading can find it through the bibliography. However, productive analyses are best conducted by trying some of these ideas, not reading about them.

a. Problem Recognition

A productive analysis has to start from the recognition that an analysis can help to solve a problem. A manager can contribute to the identification of a potential analysis by getting into the habit of asking himself, whenever he faces a new problem, whether an analyst might possibly contribute to the solution. Since there is no need for commitment to actually conduct a project at this point, the manager should set a fairly low threshold for saying "yes." As Jay (1977) notes, "the surest way not to find the best professional advice is not to look for it."

Supplementing this self-questioning, the manager should build a familiarity with the staff resources that he has available. He should investigate who his analysts are, their reputations within the organization, what they have done in the past, and their styles of operating.

From the analyst's point of view, he should continually be scouting for opportunities to contribute his analytical skills. Information for such scouting might come from asking people what they know about potential "clients" for analyses, keeping on top of external influences on the organization, and watching potential "clients" during meetings.

Importantly, a manager should involve the analyst(s) at the earliest possible stage of thinking about an analytic problem. However, this may be the most difficult thing for a manager to do. As Baker and Schaffer (1969) observe:

> "Every time staff consultants come around with their new concepts and new approaches, [the manager] runs the risk that the very foundation on which he has built his career will be eroded. He fears that the familiar body of knowledge and experience which has given him his sense of confidence may no longer be as relevant."

Similarly, Brown (1970) also recognizes one of the most serious "costs" of analysis to be "the discomfort an executive feels as he forces his traditional way of thinking into an unfamiliar mold and lays bare to [the analyst] the most delicate considerations that enter into his decision making." The first piece of advice, then, is to fight the strong psychological pressures to do otherwise and actively think of ways to use an analyst.

b. Readiness Assessment

After identifying that an analyst might contribute to the solution of a problem, the manager should assess his readiness to use advice. That is, the manager must determine whether he has the intention and the will to implement the results of the analysis when he gets them. To quote Jay (1977) again, "there is no surer (or commoner) way of wasting money on consultants than by paying for surveys, reports, and recommendations and then doing nothing about them." The same can be said for use of staff analysts, although their cost may be hidden to some degree.

The following checklist of questions (adapted from Schaffer, 1983) should help identify readiness:

1. Motivation. How strongly is the manager motivated to achieve results? How much pressure is there for improvement? How much improvement and how fast?

2. Priority. Is this the manager's most important problem? If not, how important is it?

3. Need for Analysis. Do standard procedures or routine judgments offer satisfactory resolution of the problem or is a more extensive and innovative effort needed?

4. Organization. Who in the organization really wants the analysis?

5. Attitudes. How do others feel about analysts being involved with the problem?

6. Impact. Who stands to win or lose from the results of the analysis? Did we forget about any possible loser?

7. Commitment. How willing is the manager to confront tough issues directly? To take responsibility for results?

8. Capacity for Change. What is the capacity of the manager and the organization to absorb changes that might be recommended?

Answers to these questions provide the information needed to shape a project that will produce useful results that get used. The assessment of readiness is a step that is often ignored but it is crucial to the success of the project. As Baker and Schaffer (1969) observe, "if the work is beyond what the managers are ready to tackle, or if recommended changes entail too great a sense of risk by line executives, the chances of failure and frustration are high indeed."

c. Project Formulation

After identifying a possible problem area and assessing the readiness to use the results of an analysis, a manager and analyst should meet to formulate a project. Such a formulation should consider things such as desired results, scope of the work, and whether an analysis should be done at all. Before calling in the analyst, the manager should try to describe, in as much detail as possible, the wanted end result. This helps the manager and the analyst to formulate a plan for achieving that result. Extreme care must be taken at this point to provide a description of the <u>need</u>, not the solution. As Jay (1977) puts it, "half the errors in the use of professional advisors spring from what is the equivalent of using a doctor as if he were a druggist." The manager should also be careful to identify and communicate to the analyst the real motivation for the project (even if it is a selfish reason). Otherwise, the analysis may produce the right answer to the wrong problem.

With this statement of needs, the manager and analyst can work together to devise a plan to meet the needs. This is best done with a written task statement that specifies: the reason for the project, the nature and scope of work to be done, the manager's goals, the time frame, approximate costs or level of effort, modes of working, expectations that the manager or analyst might have, and some indication of how the project will be evaluated and terminated (Steele, 1982). This task statement should provide a "road map" of the analysis but should allow for flexibility in the way that the analysis is carried out. Again, the words of Jay (1977) are appropriate: "do not trespass on the areas of professional skill and professional judgment that you hired your advisors for."

d. <u>Project Management</u>

After a project is under way, the manager should not simply sit back and wait for the results, he should participate in the review and overall management of the project. However, at the same time, the manager must resist the temptation to tell the analyst <u>what</u> to do. This can often be a difficult balance to strike, retaining control over the project while avoiding detailed direction. However, it is another key step toward ensuring that the analysis addresses the right problem.

Jay (1977) calls the non-decision of waiting for the results abdication and offers the following advice:

"You must stay in firm control of your professional so that you always know what he is doing, why, when it will be done, what its implications for and repercussions on the rest of the organization will be, and how much it is going to cost. How he is doing it is another matter. You can understand that only in the very broad terms. You are in greater danger of abdication if you have not got absolutely clear in your mind from the start precisely what the project is intended to achieve, and if you have not built the review points into it from Day 1."

Related to the subject of retaining control are the topics of "understanding" and "owernship." The skilled manager must be able to communicate with the analyst to remain in control. As Baker and Schaffer (1969) note:

"Too often, managers are inhibited in their dealings with staff specialists because they do not know the technology. As a consequence, they may fail to exploit contributions staff specialists make. Or, at the other extreme, they may expect the staff to produce miracles without their own participation or direction."

Just as the manager should avoid telling the analyst how to do his job, neither should the manager expect the analyst to do the manager's job. The manager should not abdicate his managerial responsibility and he should be involved enough in the analysis that he feels comfortable with assuming some ownership over the results. This is also good advice to the analysts. As John K. Baker, Director of Management Services at Union Carbide Corporation, explained (Baker and Schaffer, 1969):

"In every case, the object is to make certain that the management of the business has as much 'ownership' of the results of these projects as any of our staff people."

e. Project Wrap-Up

As Steele (1982) notes, "finishing a project well is as important as starting it well." Each analysis should have a clear point at which it is ended. The end point is also the time to set the stage for future interactions with analysts. Both the analyst and the manager can learn how to make future analyses more productive if this ending includes a review and evaluation of the analysis and the analysis process. If earlier suggestions were followed, a basis for evaluation will already be established. The basic evaluation question is: Did the analysis fulfill the need stated in the beginning of the project (step c) or as modified during the course of the project (step d)?

More importantly, from a learning standpoint, is an analysis of the process, the manager-analyst relationship. For example, Jay (1977) suggests asking the following questions:

> "Were there any misunderstandings? Was there information on either side that was not discovered by the other until too late? Did either side cause the other a lot of work or expense that could have been avoided? Were any shortcuts missed? Was the intitial formulation of the objective too broad, or too narrow, or too vague?"

Answers to questions such as these help form the basis for more productive analyses in the future.

f. Conclusion

The guidelines given above are offered as suggestions for improving the managerial use of staff analysts. It may sound like a lot of work to go through just to get an analysis done. However, a little work at managing the process of analysis can show big payoffs in the quality of analyses and in the quality of decisions based on those analyses.

As a final note, many managers and analysts do not follow most of these suggestions. This may explain why so many analyses end up gathering dust on shelves.

BIBLIOGRAPHY

Baker, J.K., and Schaffer, R.H. "Making staff consulting more effective." Harvard Business Review, January-February 1969, pp. 62-71.

Bobbe, R.A., and Schaffer, R.H. Mastering change: Breakthrough projects and beyond. AMA Management Bulletin No. 120. NY: American Management Associations, 1968.

Brown, R.V. "Do managers find decision theory useful?" Harvard Business Review, May-June 1970, pp. 78-89.

Fuchs, J.H. Making the most of management consulting services. NY: AMACOM Division of American Management Associations, 1978.

Jay, A. "Rate yourself as a client." Harvard Business Review, Vol. 55, No. 4, 1977, pp. 84-92.

Schaffer, R.H. "Advice to internal and external consultants: Expand your client's capacity to use your help." S.A.M. Advanced Management Journal, Vol. 41, No. 4, 1976, pp. 39-52.

Schaffer, R.H. "Launching a new consulting project." AIMC Forum, Vol. 1, No. 1, January 1983, pp. 8-10. (Publication of the Association of Internal Management Consultants, Inc., P.O. Box 472, Glastonbury, CT 06033.)

Shay, P.W. How to get the best results from management consultants. NY: Association of Consultant Management Engineers, Inc. (347 Madison Avenue, NY, NY 10017), 1974

Steele, F. Consulting for organizational change. Amherst, MA: University of Massachusetts Press, 1975.

Steele, F. The role of the internal consultant. Boston: CBI Publishing Co., 1982.

2. APPLICATION-ORIENTED DISCUSSION OF PREVIOUS RESEARCH

This chapter is devoted to a discussion of selected portions of the Wyle report. The discussion is designed to clarify and simplify the original presentation. Where appropriate, additional materials (generally consisting of suggested alternative techniques) are presented.

The topics selected for discussion are those deemed most immediately relevant to the RBS mission. Selected primarily from Chapter 1 of the Wyle report, the topics are organized within the context of three of the four areas of boating safety application identified earlier.[1] Specific topic headings are as follows:

	Section of Wyle Report in Which Discussed
2.1 Needs Assessment	
2.1.1 Methods for Detecting Clusters	
2.1.1.1 Model Forms for General Analyses	1.2.3.3
2.1.1.2 Accident Recovery Model (ARM)	1.2.5
2.1.2 Methods for Detecting Trends	New Material
2.2 Performance Prediction	
2.2.1 Moving Averages and Exponential Smoothing	New Material
2.2.2 Autoregressive Models	1.3.1
2.2.3 Forecasting Accuracy: How Much is Needed?	New Material
2.2.4 Predicting Impact	1.3.2
2.2.5 Boater Risk-Taking	1.6
2.3 Performance Evaluation	
2.3.1 Model Forms for Specific Analyses	1.2.3.2
2.3.2 Intervention Analysis	1.4.2
2.3.3 Assessment of Impact Diagrams (AID)	1.4.5

---

[1] The fourth area of boating safety application, Resource Allocation, was not addressed in the Wyle report and will be treated at length in Chapter 4.

2.1 NEEDS ASSESSMENT

Although the recreational boating fatality rate in the United States (measured in fatalities per hundred thousand boats) has declined dramatically over the past twenty years, the fact that there remain twelve to fifteen hundred fatalities each year is in itself an awesome statistic. It is reasonable to suppose that most of those fatalities could have been prevented had some missing element been present, or had some element which was present been handled differently.

A judgment as to "what went wrong?", in a given boating accident, is often evident through scrutiny of the Boating Accident Report (BAR). In other cases, a reasonable judgment might be possible through inference or knowledgeable surmise. In some fraction of cases, however, there is simply no basis for knowing what took place.

Accidents which are random and unrelated do not lend themselves to organized countermeasures. Through analysis of the BAR reports, however, one would hope to identify "patterns" which might indicate the possibility of effective corrective action. Two types of patterns are of primary interest:

a. Clusters. - Clusters are accidents or fatalities which share a common cause or a common aftermath. If the factor or factors which underlie that commonality could be identified and addressed, the number of accidents and/or fatalities might be substantially reduced. Clusters may include accidents taking place during the same time period (weekend, month, year) or in different periods.

b. Trends. - Trends are pattern changes over time. Trends in RBS accident or fatality rates are usually a tipoff that a non-random, possibly correctable process is taking place. Again, if the nature of that process could be identified, a substantial reduction in accident and/or fatality rates might be realized.

The role of the RBS analyst, then, is to search for clusters or trends of sufficient magnitude to indicate (a) the need for, and (b) the possibility of, effective corrective action.

2.1.1 Methods for Detecting Clusters

The Wyle report presents two sets of methods useful in the detection of clusters. The first set of methods is discussed under the rubric, Model Forms for General Analyses, and is found in Section 1.2.3.3; the second set is termed the Accident Recovery Model (ARM) and is found in Section 1.2.5. The discussion which follows addresses each of these topics in turn.

2.1.1.1 Model Forms for General Analyses

These are models which make use of computerized methods to provide counts of the number of accidents in the data base which present certain commonalities. Each of the models is of the form known as a "decision tree"; i.e., the model consists of a series of logical branches leading from (1) what is contained in the accident report to (2) what is likely to have been the factor or factors involved either in causing the accident or in preventing a successful recovery. Using these models, the analyst is able to group accidents by failure cause and/or reason for unsuccessful recovery, and to obtain counts for each such grouping.

Of the models suggested by Wyle for use in this form of analysis, the following are the most relevant:

a. Cause Assessment Tree. - This is a format used to assign a primary cause to each accident (or to each boat in a two-boat accident). The nodes of the decision tree must be properly ordered to assure that the appropriate accident cause is selected. Figure 1-3 of the Wyle report illustrates a tree of this form; Figure 1-4 is the coding form used in its application.

b. Component Tree Model. - This is a variation of the Cause Assessment Tree in which several trees, each covering a different aspect of the accident (or recovery), are used. The branches within each tree are mutually exclusive; aspects of the accident/recovery which are not mutually exclusive are handled in different trees. (The Accident Recovery Model, shortly to be discussed, makes use of this model form.)

c. Fault Tree. - This is a decision tree which focuses on the logical relationship among the various elements which could, in theory, lead to system failure, i.e., to an accident or unsuccessful recovery. To use this technique, the analyst must be able to assign probabilities of occurrence to each separate failure mode. The absence of useful data on the various failure probabilities involved in boating accidents limits the usefulness of this technique, generally speaking, to accidents due to hardware failure. In the words of the Wyle report, "the fault tree is generally not an appropriate model form for modeling recreational boating accidents."

Whichever of these model forms is used, its primary purpose is to permit the analyst to recognize (and to count) accidents which have one or more items in common. The primary value of these models is in identifying areas of boating safety which might require corrective action ("needs assessment"). Decision trees, similar to these but of a more directed nature, are useful in other contexts as well: whether the analyst is predicting the impact of a proposed new standard ("performance prediction") or evaluating the impact of an existing standard ("performance evaluation"), he clearly must know the frequency with which accidents and/or fatalities relating explicitly to that standard have occurred in the past. Models of this explicit nature are termed Model Forms for Specific Analyses and are discussed in Section 2.3, Performance Evaluation.

2.1.1.2 Accident Recovery Model (ARM)

ARM is a particular accident profile model which incorporates the general principles of the Component Tree Model, described above. The model was developed by Wyle Laboratories for the purpose of studying variations in survival probability as a function of various factors. The notion underlying the model is to use a sample of accident data to arrive at inferences concerning the universe of accidents and recoveries as a whole. To do this, the model makes use of statistical weights to offset the bias which might result if the sample were not totally representative of the universe.

The bias removal feature is perhaps the most noteworthy feature of the ARM Model. The Wyle report divides the universe of boating accidents into a 6 x 6 matrix: Boat Type by Accident Type. The six boat types are:

- o Open manual
- o Open power
- o Cabin motorboat/houseboat
- o Sail/auxiliary sail
- o Canoe/kayak
- o Other

and the six accident types are:

- o Capsizing/swamping
- o Collision/grounding
- o Fall overboard
- o Fire/explosion
- o Hit by boat or prop
- o Other

There are, as one would expect, a different number of accidents in each of the 36 cells of the matrix. Because of this, it is impossible to arrive at a sample which is equally representative, i.e., has the same sampling rate, in each and every cell. For example, if Cell A (hypothetical) had 100 accidents, Cell B had 10, and Cell C had 2, a 10% across-the-board sampling rate would result in 10 accidents from Cell A, 1 accident from Cell B, and 0.2 accidents from Cell C. There is no way, however, to include two-tenths of an accident in a sample; the sample either includes the accident or it does not. To include exactly one accident from Cell C, however, would mean increasing the sampling rate for that cell to 50%, a sampling rate which would be totally inappropriate (and costly) for Cell A.

The obvious solution is to let the chips fall where they may -- to use a different sampling rate for each cell. Ordinarily, one would wish to oversample the smaller cells and undersample the larger ones. Once this is done, however, the data from the undersampled and oversampled cells cannot simply be combined, for to do so would bias the overall results. As the Wyle report points out, the data in each cell must be weighted to bring each cell back to parity with the others. Parity is 100; thus, a cell which has a sampling rate of 20% must be weighted by a factor of 5, a cell which has a sampling rate of 12.5% must be weighted by a factor of 8, and so on. In brief, the greater the sampling rate -- i.e., the greater the percentage of accidents sampled in a given cell -- the smaller the weight assigned to that cell when merging the results.

The ARM Model actually uses two sets of weights -- one for fatalities and one for recoveries. The weights used for fatalities are simply the ratio of the number of fatalities in the universe to the number of fatalities in the sample. Thus, if a given cell accounted for 120 fatalities in the total universe but only 10 in the sample, the fatality weight for that cell would be 12. This means that in studying the fatalities associated with that cell, all characteristics of the sample would be weighted by a factor of 12. Other cells would of course receive different weights.

Weights for studying recoveries are more difficult to define. Since the purpose of the model is to study variations in survival probability, it is important that all opportunities for recovery in a given accident be counted. Not all accidents are identical in this respect, however. In cases involving capsizing/swamping, collision/grounding, and fire/explosion, all passengers on board are potential victims; thus, all who survived should be counted as having experienced a successful recovery. In other cases, however, only a single or at most a limited number of passengers were potential victims; in those cases, only that passenger (or limited number of passengers) should be counted. The rationale used in the ARM Model for counting recoveries, making use of information contained in the accident report, is as follows:

a. For all accidents other than falls overboard, hits by boat or prop, and "other", Wyle defined "the number of persons recovered" as the total number on board minus those who failed to survive. Both of these items are included in the accident report.

b. For hits by boat or prop, Wyle assumed there was only one potential victim per accident. Thus, "the number of persons recovered" was defined as the number of non-fatal accidents of this type. Again, this information is available from the accident reports.

c. For falls overboard and "other", Wyle was initially inclined to count only one recovery per accident. The data disclosed, however, that there were more fatalities of this type than there were boats, indicating that there was often more than one person at risk per accident. Since there was no way of deriving the exact number of potential victims from the accident reports, Wyle applied the same recovery weights as in the first case, i.e., "number of persons recovered" was defined as the total number of persons on board minus fatalities.

Using these procedures to derive both weights and survival probabilities, Wyle applied the ARM Model to a sample of Coast Guard accident reports for the year 1975. The sample selected covered a greater percentage of fatalities (277 out of 1,489) than it did of recoveries (1,229 out of 18,318). Major findings cited were as follows:

- o Victims from boats with sufficient PFDs had a much greater probability of survival than those from boats lacking in PFDs.

- o The probability of recovery increased with:

  - increasing people on board
  - increasing boat length
  - increasing water temperature
  - decreasing distance to shore or another vessel.

- o Victims from canoes, kayaks, open manual boats, and "other" boats had significantly lower chances for survival than victims from powerboats, cabin motorboats, houseboats and sailboats.

- o Manually powered boats led to an unusually low probability of recovery for victims in reported accidents.

- o Victims from reported collisions, groundings, fires and explosions fared well, while victims in reported capsizings/swampings and falls overboard had much reduced chances for survival. Hit-by-boat-or-prop and "other" victims in reported accidents had intermediate probabilities of recovery (approximately 0.89).

These conclusions, taken verbatim from the Wyle report, illustrate the types of assessments possible using the techniques described in this section. The reader is cautioned, however: these assessments are only as valid as the data are complete. One of the reasons why "victims from reported collisions, groundings, fires and explosions fared well" is that a presumably large percentage of such accidents are reported whether or not there was a fatality. Thus, a large number of recoveries, as well as of fatalities, are included in the data base and the calculated survival probability is relatively high. On the other hand, the reason why "victims in reported capsizings/swampings and falls overboard had much reduced chances of survival" may well be that a large percentage of those accidents are not reported unless there was a fatality. Thus, the calculated survival probability for these types of accidents may well be understated. Wyle Laboratories shows its sensitivity to this issue by its inclusion of the word "reported" in its conclusions.

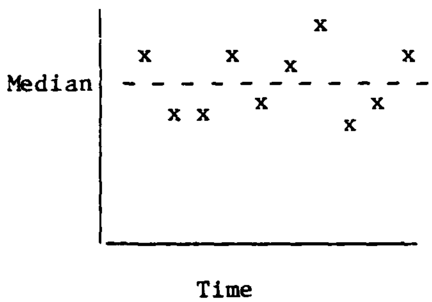
2.1.2 Methods for Detecting Trends

Another essential component of the analyst's job is the early detection of trends. Trends, as noted earlier, are usually a tipoff that a non-random, possibly correctable process is taking place. Example: Suppose it became evident, after years of relative stability, that the fatality rate involving non-use of personal flotation devices (PFDs) had begun to rise. In these circumstances, one might conclude that an educational campaign directed toward re-emphasizing the use of PFDs might be in order. At the very least, an investigation of the reasons for the rise -- possibly using some of the models described earlier -- might be initiated.

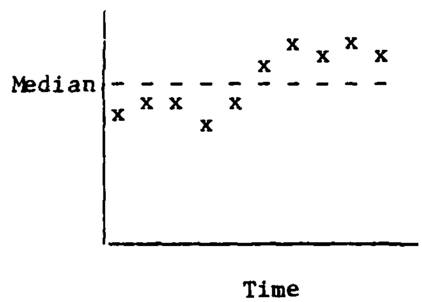
A conventional method for detecting the presence of a trend is to calculate a least squares regression equation and test the slope against zero. If the slope differs significantly from zero, a trend may reasonably be judged to exist. Readers unfamiliar with the use of regression methods may find the McGraw-Hill publication, Statistical Package for the Social Sciences, useful reading; particular attention should be paid to the section entitled Statistical Inference in Regression Problems, in which the methods used to test regression slopes for significance are discussed. Another useful reference is the chapter on Multiple Linear Regression contained in the BMD (Biomedical Computer Program) Manual, available from the University of California School of Medicine.

Those who lack ready access to a computer, or who believe conventional regression methods are inappropriate for their application, may prefer alternative methods. One such method is based on the theory of runs. The theory states, in essence, that in any time series consisting of purely random elements, there should be a sprinkling of observations both above and below the median, whereas if the series were non-random, observations above and below the median should tend to cluster. The distinction between these two phenomena is illustrated below:

a. Random



b. Non-Random



Each uninterrupted sequence of observations *above or below the median* constitutes a single run. In the random situation shown above, there were a total of seven such runs: first an "above," followed by two "belows," then an "above," a "below," two "aboves," two "belows," and finally, an "above." In the non-random situation, there were only two runs: five "belows" followed by five "aboves."

Although the literature on the theory of runs is not extensive, the subject has received attention in industrial quality control applications where the timely detection of trends can be quite important. If the number of runs above and below the median, in a given set of observations, tends to be smaller than one might expect, one might conclude that the process is non-random and that action is needed to bring it back into control. Table 2.1 (next page) lists the minimum number of runs, r, which one would expect in a series of n observations. These numbers have the following meaning: if the number of runs above and below the median, in a series of n observations, is r or less, that number is said to be statistically significant, i.e., the chances are less than 5% that the series in question is random.

Table 2.1. Critical Number of Runs Above and Below the Median in a Series of n Observations

Number of Observations (n)	Critical Number of Runs (r)
10	2
12	3
14	3
16	4
18	5
20	5
22	6
24	7
26	8
28	9
30	10
32	11
34	11
36	12
38	13
40	14

Adapted from Swed and Eisenhart, "Tables for Testing Randomness of Grouping in a Sequence of Alternatives," Annals of Mathematical Statistics, Vol. 14, 1943, pp. 83-86.

To illustrate the use of the table: The critical number of runs for a set of 24 observations is seen to be 7. This means that if a series of 24 observations falls in such a manner that there are only seven runs above and below the median, the analyst is justified in concluding that the series is non-random and may require action. If there were eight runs or more, no such conclusion would be justified (at the 95% level of significance).

ILLUSTRATION:[1]

(a) For the past 24 months, the seasonally adjusted number of fatalities per 100,000 boats involving non-PFD users has been as follows:

14,13,17,16,19,17,18,15,20,22,19,20,23,24,21,19,23,22,25,21,22,24,26,23

Does the series appear to be random?

(b) How about this series of observations?

14,20,22,13,17,16,19,20,23,19,17,24,21,19,25,26,21,22,18,15,23,22,24,23

---

[1] NOTE: The numbers presented in this illustration are hypothetical.

SOLUTION:

(a) An examination of the series shown on the preceding page shows that exactly half of the observations (12) have a value of 20 or less; half have a value of 21 or greater. The median is therefore midway between 20 and 21, i.e., 20.5. Labeling the observations below the median by "B" and those above the median by "A," we have the following sequence of values:

B,B,B,B,B,B,B,B,B,B, A, B,B, A,A,A, B, A,A,A,A,A,A,A,A

Each unbroken underscore represents a single run. There are a total of six such runs. Since this number is less than the critical value shown in Table 2.1, the series is judged to be non-random. Apparently fatalities due to non-use of PFDs are on the rise.

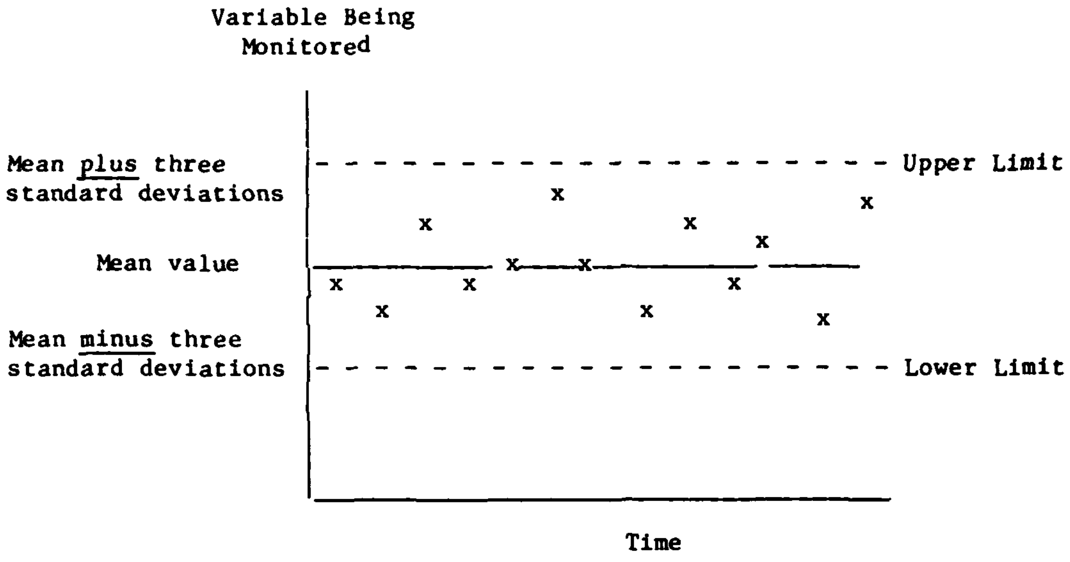
(b) These are the same 24 observations, in a different time sequence.

Again the median value is 20.5. This time, however, the series of A's and B's is as follows:

B,B, A, B,B,B,B,B, A, B,B, A,A, B, A,A,A,A,A, B,B, A,A,A,A

The number of runs, in this instance, is ten. Since this number exceeds the critical value shown in Table 2.1, the observed series does not appear to represent a trend.

Another set of methods, drawn from the industrial quality control milieu, which the analyst may find helpful is the use of "control" charts. These are charts on which the analyst periodically plots the performance associated with a given process, comparing the numbers derived against predetermined upper and lower control limits. The upper and lower limits are normally defined as the mean value of the process plus and minus some number of standard deviations (usually three). A typical control chart is shown below:



When an observation falls either above the upper or below the lower control limit, the analyst is reasonably justified in concluding that an "out of control" event has taken place. In this sense, control charts are more responsive to "shifts" or "excursions" from an established pattern than are either of the methods discussed earlier. Both regression methods and runs above and below the median are more suited to detecting gradual shifts over time.

The reader interested in learning more about control charts may wish to read E. L. Grant, Statistical Quality Control, McGraw-Hill. Chapters IV and V, Why the Control Chart Works, are particularly relevant.

2.2 PERFORMANCE PREDICTION

Once a given safety need or set of needs has been identified, the countermeasures available to meet those needs can take on a variety of forms. As noted in the Introduction, these activities range from safety standards to educational campaigns to broad initiatives designed to create a safer boating environment.

Some of these measures are adopted based on pure judgment -- because they seem a sensible, cost-effective thing to do. Others, however -- particularly those which involve a heavy investment in resources -- require careful study prior to implementation. Before a potentially controversial or costly measure can be undertaken, it must be studied in sufficient depth to assure that the effort involved will be productive and that the benefits are likely to exceed the cost. Where several alternatives are being considered, prudent choices must be made based on likely benefits and likely cost.

This section focuses on the prediction of benefits. Benefits are, as noted earlier, normally expressed in terms of reduced accidents, property damage, and/or fatalities. Whichever of these impact variables is employed, the first step in the prediction process is to forecast what the value of that variable would be if the countermeasure in question were <u>not</u> to be adopted, i.e., if the status quo were to be maintained. The difference between the forecast value and the predicted <u>modified</u> value due to the proposed intervention is then the estimated measure of impact.

The discussion in this section centers initially on the forecasting process, then on methods for estimating impact. A discussion of boater risk-taking, i.e., the tendency for boaters to take greater risks when presented with seemingly safer equipment or a "safer" environment, is also included. As the title of this section implies, the discussion centers on the prediction process. How to pick and choose from among competing possibilities once all the predictions have been made constitutes the subject of Chapters 3 and 4 on multiattribute utility analysis and resource allocation methods respectively.

2.2.1 Forecasting Made Simple: Moving Averages and Exponential Smoothing

In seeking to extrapolate the past, one should not be misled by momentary valleys or peaks. Suppose, for example, that for the past five years the number of fatalities (per 100,000 boats) resulting from a given accident cause has been as follows:

42, 59, 47 62, 71

If asked to estimate the fatality rate currently associated with the given accident cause, the analyst might choose to reply "71," the figure for the most recent year. This reply, however, ignores the evidence of prior years; thus, if 71 happens to be a momentary peak, the analyst will have overstated the case.

An alternative to using the most recent figure would be to express an average over all five years of data. In this case, that average would be

$$\frac{42 + 59 + 47 + 62 + 71}{5} = 56.2$$

If the data are trending, however, this value might well be an understatement. The analyst might therefore choose to compromise by selecting only the last three (or four) years. Smoothing in this manner -- by presenting the average values for the last m time periods (years, months, weekends) -- is known as smoothing through the use of moving averages.

Exponential smoothing is another possibility. Unlike moving averages, in which all years are assigned equal weight, exponential smoothing assigns less weight to prior years' observations than to those which are current. It does this through the following formulation:

$$Y_t = Sy_t + (1-S)Y_{t-1}$$

where Y_t = the smoothed value of y at time t

Y_{t-1} = the smoothed value of y at time t-1

y_t = the actual value of y at time t

S = a smoothing constant between 0 and 1

Setting S, the smoothing constant, close to 1 causes the present value of y to predominate; setting S close to 0 increases the importance of past values. No matter what the value of S, however, the present value of y always receives more weight than those preceding it.[1] Exponential smoothing therefore differs from other methods of smoothing, such as moving averages, in which all past values (up to a certain point) receive equal weight. The smoothing constant S bears the following relationship to the number of years (m) used in calculating a moving average:

$$S = \frac{2}{m + 1}$$

That is to say, a smoothing constant of 0.5 is essentially equivalent to a three-year moving average (m=3) in terms of the amount of past history retained; a smoothing constant of 0.4 is tantamount to a four-year moving average (m=4) and so on. The following table shows the rough equivalence between the smoothing constant S and the number of years in the moving average, m.

Table 2.2. Equivalence Between m and S

Number of Years in Moving Average (m)	Analogous Smoothing Constant (S)
3	0.50
4	0.40
5	0.33
6	0.29
7	0.25
8	0.22
9	0.20
10	0.18

[1] To see this, note that the current smoothed value, Y_t, can be rewritten as an infinite series, i.e.,

$$Y_t = Sy_t + S(1-S)y_{t-1} + S(1-S)^2y_{t-2} + \ldots$$

This representation shows that y_t always receives more weight than y_{t-1} since S is always greater than S(1-S). Similarly, y_{t-1} always receives more weight than y_{t-2}, and so on.

A basic discussion of exponential smoothing techniques may be found in Barry Shore, Operations Management, McGraw-Hill, 1973, Chapter 11. A more advanced treatment is contained in Robert G. Brown, Smoothing, Forecasting, and Prediction of Discrete Time Series, Prentice-Hall, 1963.

ILLUSTRATION:

Given the series of fatalities per 100,000 boats presented earlier in this section:

42, 59, 47, 62, 71

calculate a "smoothed" value for the most recent year using smoothing constants of 0.5, 0.4, and 0.33 respectively. (Assume that the first year's value, 42, has already been smoothed.) Compare these values against those obtained through three-, four-, and five-year moving averages.

SOLUTION:

In the case of S equal to 0.5, the series of smoothed observations is calculated as follows:

First year: $Y_1 = 42$

Second year: $Y_2 = 0.5(59) + 0.5(42) = 50.5$

Third year: $Y_3 = 0.5(47) + 0.5(50.5) = 48.8$

Fourth year: $Y_4 = 0.5(62) + 0.5(48.8) = 55.4$

Fifth year: $Y_5 = 0.5(71) + 0.5(55.4) = 63.2$

The most recent figure, 63.2, is the current smoothed estimate. Although substantially lower than the most recent observation (71), it is still somewhat higher than any of the following estimates based on moving averages:

Three-year moving average = (47 + 62 + 71)/3 = 60.0

Four-year moving average = (59 + 47 + 62 +71)/4 = 59.75

Five-year moving average = (42 + 59 + 47 + 62 + 71)/5 = 56.2

The fact that exponential smoothing, which weights the more recent observations more heavily, produces higher estimates than moving averages is a sign that the data may be trending upward.

SOLUTION (cont'd):

Similar calculations for S equal to 0.4 and 0.33 lead to the following sequences of exponentially smoothed observations:

S = 0.4

First year:	$Y_1 = 42$	
Second year:	$Y_2 = 0.4(59) + 0.6(42)$	= 48.8
Third year:	$Y_3 = 0.4(47) + 0.6(48.8)$	= 48.1
Fourth year:	$Y_4 = 0.4(62) + 0.6(48.1)$	= 53.7
Fifth year:	$Y_5 = 0.4(71) + 0.6(53.7)$	= 60.6 (current estimate)

S = 0.33

First year:	$Y_1 = 42$	
Second year:	$Y_2 = 0.33(59) + 0.67(42)$	= 47.6
Third year:	$Y_3 = 0.33(47) + 0.67(47.6)$	= 47.4
Fourth year:	$Y_4 = 0.33(62) + 0.67(47.4)$	= 52.2
Fifth year:	$Y_5 = 0.33(71) + 0.67(52.2)$	= 58.4 (current estimate)

Note that as S gets smaller, i.e., as the emphasis shifts more heavily from the present to the past, the current estimate declines (from 63.2 to 60.6 to 58.4). Again, this is a sign that the data are trending upward.

2.2.2 Autoregressive Forecasting Models

The simplicity of the methods just described may also be their major drawback: they smoothe but do not directly predict. They remove the unevenness of the past but provide no formulation for the future.

Methods which provide such a formulation include traditional regression models (discussed in the preceding section) and a relatively new set of techniques, discussed in Section 1.3.1 of the Wyle report, known as Box-Jenkins ARIMA (Autogressive Integrated Moving Average) models.

For an understanding of the ARIMA models, it is necessary first to understand the backshift operator B. B is simply a symbol which defines the following operation:

$$By_t = y_{t-1}$$
$$By_{t-1} = y_{t-2}$$
$$\vdots \qquad \vdots$$

where y_t is the observed value of the variable of interest at time t.

The operator B thus shifts the variable to which it is applied one unit backward in time. If time is measured in months, successive monthly observations can be simply expressed as follows:

$$By_{June} = y_{May}$$
$$By_{May} = y_{Apr}$$

and so on.[1]

The operator B can also be used to express successive differences among observations, as shown below:

$$(1-B)y_{June} = y_{June} - y_{May}$$
$$(1-B)y_{May} = y_{May} - y_{Apr}$$
$$\vdots \qquad \vdots$$

This use of differences, rather than raw observations, is particularly relevant when using ARIMA models. The reason is that the raw data often fail to display a constant mean and variance over time. The latter property, known as stationarity, is essential to the use of these models. Fortunately, if the observations themselves are not stationary, it frequently happens that their differences are -- or, if not the first differences, then the second. Whichever set of differences first satisfies the condition of stationarity normally governs the form of ARIMA model used.

---

[1] Weeks -- or more precisely, weekends -- are the most consistent time periods for tabulation and analysis of boating accidents and fatalities. This is a recent finding, not recognized at the time the Wyle report was drafted. For simplicity, however, the discussion here remains in terms of months.

ILLUSTRATION:

Given the following series of observations:

Month =	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
y =	27	33	75	133	196	199	192	147	90	66	50	28

calculate the associated first and second differences. Comment on their apparent stationarity.

SOLUTION:

The first differences for the first two months are as follows:

$$(1-B)y_{Feb} = y_{Feb} - y_{Jan} = 33 - 27 = 6$$

$$(1-B)y_{Mar} = y_{Mar} - y_{Feb} = 75 - 33 = 42$$

The remaining first differences, in order, are:

58, 63, 3, -7, -45, -57, -24, -16, -22

These differences are clearly non-stationary, i.e., they do not appear to represent random variations about a constant mean. The first five values (6, 42, 58, 63, and 3) all lie above the median and the remaining five below. Thus, there are only two runs above and below the median in a series of ten observations. This finding, according to Table 2.1 on page 23, clearly indicates a non-random process.

The <u>second</u> differences are simply the differences among the first differences. They are in order:

36 (i.e., 42 minus 6), 16, 5, -60, -10, -38, -12, 33, 8, -6

Of these differences, five are positive and five negative and they are well intermingled (4 runs in a series of 10 observations), pointing toward the likelihood of a stationary process.

The values in this time series were taken from Table 1-8 of the Wyle report. They represent monthly forecasts of boating fatalities for 1977, derived by Wyle using the ARIMA model.

Seasonality

Seasonality is another feature which the ARIMA models are designed to accommodate. In the same way that B denotes a shift of one time unit, the symbol B^i denotes a shift of i time units. In particular, if time is measured in months, B^{12} denotes a shift of exactly twelve months. This notation permits successive yearly differences to be expressed as follows:

$$\begin{aligned}(1-B^{12})y_{June\ 81} &= y_{June\ 81} - y_{June\ 80}\\ (1-B^{12})y_{May\ 81} &= y_{May\ 81} - y_{May\ 80}\\ &\vdots\end{aligned}$$

While it is true that the taking of yearly differences avoids seasonality, the resulting series may nonetheless not be stationary. If so, taking the differences among these differences may be all that is needed to attain stationarity. Typically, such a series, based on second differences, would be expressed as follows:

$$\begin{aligned}(1-B)(1-B^{12})y_{June\ 81} &= [y_{June\ 81} - y_{June\ 80}] - [y_{June\ 80} - y_{June\ 79}]\\ (1-B)(1-B^{12})y_{May\ 81} &= [y_{May\ 81} - y_{May\ 80}] - [y_{May\ 80} - y_{May\ 79}]\\ &\vdots\end{aligned}$$

Generally, by the time second differences are reached, most series will have attained stationarity (see exercise on preceding page).

Box-Jenkins ARIMA models can be expressed in many ways using different backshift combinations. The specific model used in the Wyle report is of the form:

$$(1-B)(1-B^{12})y_t = (1-\theta_1 B)(1-\theta'_1 B^{12})a^t$$

where y_t, as before, denotes the observed value of the variable of interest at time t, θ_1 and θ'_1 are parameters whose values are determined so as to produce the best fit to the available data, and a_t is the error term (difference between actual and predicted value) associated with time t.

Comparison of Model Results

To investigate the suitability of the selected ARIMA model, Wyle compared its performance to that of a power function regression model of the form

$$y = a(x-x_o)^b$$

where y is the chosen impact variable (number of accidents, fatalities, etc.), x is the year for which a prediction is sought, x_o is an arbitrarily defined base year, and a and b are parameters whose numerical values are determined so as to produce the best "fit."

In performing this comparison, Wyle applied the regression model to the number of yearly fatalities starting in 1960, and the ARIMA model to the number of monthly fatalities starting in January 1969. Both models were then used to "forecast" values for the years 1975 through 1977. Of the two models, the ARIMA model did better, i.e., yielded errors which were only a third to a fourth as large as those yielded by the regression model. Table 2.3 summarizes these results:

Table 2.3. Comparison of ARIMA and Power Function Regression Model Prediction Errors

Forecast Based on Data Through Calendar Year	Year Being Forecast	Prediction Error (Forecast minus Actual) Regression Model	ARIMA Model
1974	1975	109	-40
	1976	333	143
	1977	306	-72
1975	1976	316	81
	1977	288	-12
1976	1977	239	-84

As shown in Table 2.3, the regression model led to consistent overestimates, sometimes as much as 25% off the mark[1], whereas the ARIMA model yielded both over- and under-estimates and was generally accurate to within five or six per cent. It should not be concluded, however, that ARIMA models are necessarily superior in all settings: this particular comparison was biased in the ARIMA model's favor through the choice of start date -- 1960 in the case of the regression model, 1969 in the case of ARIMA. Since the 1960's saw a gradual rise in boating fatalities, followed by a decline starting in 1971, a regression model based on a start date of 1960 is at a disadvantage. As a result of over ten years of initially rising data, the model failed to follow the 1971 downturn, thereby yielding consistently high forecasts in the years ahead. The problem is thus a fault not only of the model but of its application: two separate regression equations should have been used, one for the years of rising data and one for the years of decline.

One of the advantages of the ARIMA models is the fact that they are responsive, by and large, to the data. They have two major disadvantages, however. Their workings are difficult, particularly for the uninitiated, to understand and explain. Secondly, they can be expensive to run. Before deciding to employ such models, therefore, the analyst is advised to ask: do I really need the greater accuracy associated with the use of mathematically sophisticated methods? Could acceptable results not be achieved using one of the following simpler approaches?

a. A straightforward regression model, linear or otherwise, matched to the general character of the data.

b. Exponential (or other) smoothing, followed by forecasts based on expert judgment.

The value of combining smoothing techniques with subjective judgment should not be minimized. If smoothing had been applied to the fatality data cited in

---

[1] The actual number of annual fatalities between 1975 and 1977 ranged between thirteen and fourteen hundred.

the Wyle report, the analyst would have had no difficulty in recognizing that fatalities peaked in the early 1970's and that the forecast for 1977 should have been substantially lower than the 1,551 predicted by the regression model. For example, had exponential smoothing been used with a smoothing constant of 0.5, the following smoothed values would have resulted for the years 1971 through 1975:

1971 - 1,480
1972 - 1,459
1973 - 1,606
1974 - 1,526
1975 - 1,496

The analyst, observing a decline of 110 fatalities between 1973 and 1975, might well have predicted a similar decline between 1975 and 1977, resulting in a forecast of 1,386 for the latter year. The actual number of fatalities reported that year was 1,312; an error of this magnitude (minus 74, or about 5%) would have been essentially the same as that shown for the ARIMA model in Table 2.2.

Moving-average smoothing would have done equally well. If the analyst had smoothed through the use of three-year moving averages, the resulting smoothed values would have been

1971 - 1,479
1972 - 1,591
1973 - 1,546
1974 - 1,555
1975 - 1,392

Again, it is likely the analyst would have reached a forecast in the thirteen hundreds for 1977.

The preceding discussion is admittedly speculative and after the fact. One of the advantages of advanced, sophisticated forecast models is that they minimize the need for subjective judgment and lessen the possibility of self-serving misapplication. In forecasting as in other applications, however, subjective judgment cannot be avoided, i.e., ultimately plays a role. Rigidly formal analytic methods are not an end in themselves; their role is to guide, not to bind. If the use of formal methods leads to results which are inconsistent with reason and common sense -- if a regression equation shows continued rises when the last five years have shown declines -- the analyst is justified in questioning the results and either modifying them or selecting another model.

2.2.3 Forecast Accuracy: How Much Is Really Needed?

In forecasting boating fatalities as a preliminary to predicting impact, the question of how much accuracy is needed depends on the subsequent application.

EXAMPLE: The predicted number of recreational boating fatalities in 1977, according to the Wyle report (p. 1-57), was 1,551 based on the power function regression model and 1,236 based on the ARIMA model. Suppose, hypothetically, these alternative forecasts were applied to three separate countermeasures being considered for adoption. If Countermeasure A is estimated to reduce fatalities by 4%, Countermeasure B by 6%, and Countermeasure C by 8%, it follows that the use of those forecasts would have resulted in the following predicted number of lives saved:

	Number of lives saved in 1977 according to:	
	Regression Model	ARIMA
Countermeasure A	62	49
Countermeasure B	93	74
Countermeasure C	124	99

If the goal of the analyst is simply to rank the countermeasures according to the single measure "number of lives saved", it matters little which forecast is used; in either case, C enjoys a substantial superiority over B, and B over A. If, however, the goal of the analyst is to trade-off performance versus cost, it may be important to distinguish between a marginal superiority of 31 lives per year (as predicted based on the regression model) versus one of only 25 (based on ARIMA).

The point made above, concerning the prediction of fatalities, applies as well to any of the other criterion measures (accidents and property damage) which might be used. Before any investment is made in sophisticated forecasting methods, the analyst should examine the decision context to determine if micrometer-like precision is really needed or whether less precise methods, properly applied, might not be adequate.

The analyst can often defend against forecasting errors by simply conducting sensitivity analyses, i.e., by applying several different forecasts, within a reasonable range of values, to each of the impact variables involved. If the decision reached is unaffected by the forecast used, forecasting accuracy is clearly not a factor. If, however, the decision changes as the forecast changes, the analyst and/or decision-maker must decide which of the alternative forecasts more nearly approximates the truth. For example, if the "break" point (the point at which two proposed countermeasures provide essentially equal benefits) is 1,400 lives per year, a judgment need simply be made as to whether the "correct" forecast is greater than or less than 1,400. Judgments of this nature can often be made with no recourse to advanced mathematical techniques.

2.2.4 Estimating Impact

Once a suitable impact variable has been agreed upon and an appropriate forecast (or set of forecasts) of that variable has been reached, the next step is to predict the amount by which those forecasts are likely to change if the proposed countermeasure is adopted. The magnitude of that change (measured in numbers of lives saved, accidents avoided, etc.) is then the estimated measure of impact.

The calculations involved in this process are relatively straightforward. One such process is outlined in the Wyle report in Section 1.3.2, Predicting Benefits of New Boat Standards. A more generalized version of that procedure is illustrated below. As in the Wyle report, the discussion uses "fatalities" as the impact variable.

a. The first step is to identify which segments of the boating accident universe are likely to be affected by the proposed countermeasure. Depending on the nature of the countermeasure, only certain boat types or certain accident types may be affected.

b. The second step is to ask if all segments of the universe which <u>are</u> affected will be affected equally, or whether some segments are more likely to be affected than others. The Wyle report suggests one such possibility: Age of Boat is important when new boat standards are being considered. Depending on the nature of the countermeasure, other variables which might be expected to display differential impact include: Accident Type (load-related fatalities, for example, as opposed to those resulting from swamping or flooding), Time of Day (nighttime fatalities as opposed to those occurring in the day), etc.

c. The third step is to establish what proportion of the fatality rate has traditionally been associated with each of the accident groupings identified in step "b". The disaggregation performed at this point need not be overly detailed; it need simply correspond to the specific groupings identified in the previous step; i.e., Age of Boat if age is important, Time of Day if time of day is important, and so on. For example, if "capsizing" fatalities, "swamping" fatalities, and "all others" are judged to be the only groupings likely to display different impacts, the analyst need only focus on those three categories. In 1982, the percentage of fatalities accounted for by each of these categories was as follows:

	Percentage of All Fatalities
Capsizing	34.7%
Swamping/flooding	7.3%
All others	58.0%

d. The fourth and final step is to multiply each of the percentages just derived by the predicted <u>percentage</u> reduction in fatalities for accidents of that type. For example, if a separate accident analysis indicated that capsizing fatalities are likely to be reduced by 10%, swamping/flooding fatalities by 15%, and all other fatalities by 5%, the overall <u>weighted average reduction in fatalities</u> is simply

$$(10\% \times 34.7\%) + (15\% \times 7.3\%) + (5\% \times 58.0\%) = 7.465\%$$

The latter percentage, applied to the predicted number of fatalities in any given year, constitutes the estimated number of lives saved in that year. For example, if the predicted number of fatalities is 1,250, the predicted number of lives saved that year is 1,250 X 7.465%, or 93.

2.2.5 Boater Risk-Taking

One of the factors which complicates the estimation of impact is the uncertainty associated with boater risk-taking: Will the boater, placed in what appears to be a safer environment, take risks he might previously have avoided? If so, fatalities may not actually be reduced or, if they are, the reduction may be less than expected.

The Wyle report addresses this problem by suggesting several methods for predicting boater behavior. One method (Section 1.6.3) involves a survey; the other (Section 1.6.4) an experimental design. Both methods are based on the notion that there exists a continuum along which the degree of boating hazard can be expressed and that boaters can be evaluated in terms of where along that continuum they are likely to decide to remain ashore rather than risk life and property. Section 1.6.2 of the Wyle report discusses the development of such a scale.

Whatever the theoretical merits of the methods suggested by Wyle, they present several practical difficulties, not the least of which is cost. The use of these methods, therefore, is not encouraged unless:

a. The issue of boater behavior is vital to the decision process, i.e., no decision can be reached until it is resolved.

b. Expert judgment and/or analysis of prior accident data offer no hope of a reasonable answer.

c. The hypothetical questions asked (if it is a survey) and the experimental conditions observed (if it is an experiment) can be couched in terms which are reasonably realistic.

In a survey or an experiment of this nature, realism can be hard to attain. The problems associated with survey questions are obvious: respondents tend to provide cosmetically acceptable replies, then behave as they choose. Experimental methods, although more costly, at least have the potential for providing important insights. Even with an experiment, however, there are many factors, difficult to evaluate, which can obfuscate the results. These include:

(1) Skill of the boater. -- Even if it were possible, say, to determine that 20% of the boating population are now likely to take risks they might previously have avoided, this knowledge in itself is insufficient. It is equally important to know whether that 20% represents boaters of all skills, representatively distributed, or whether it consists primarily of skilled boaters perfectly capable of handling the greater risk.

(2) Number, age, and type of boating companions. -- Decisions to launch or not to launch are obviously affected by who else is along. Boaters who deliberately court a risk when accompanied by skilled companions might well choose to remain ashore if it meant jeopardizing the lives of children.

(3) Boater's other options. -- Decisions which a boater makes when he has limited alternatives, might well go the other way if he had other, equally attractive options on that given day.

Each of these factors complicates enormously any data gathering effort devoted to measuring boater behavior. Item (1) -- boater skill -- affects the interpretation of the data and must somehow be measured. Items (2) and (3) are typical of the many intervening variables which can affect boater behavior on any given day; in a large sample, these factors tend to balance out, but in a small sample, they can lead to unacceptable bias.

2.3 PERFORMANCE EVALUATION

Boating safety analysts are often called upon to evaluate the impact of safety standards, regulations, and/or educational programs which have been operational for some time ("Performance Evaluation"). Three sets of methods useful in this regard are discussed in this section:

a. The first set of methods, Model Forms for Specific Analyses, are a variant of the model forms for general analyses previously described in Section 2.1, Needs Assessment. Unlike the earlier "general" model forms, which attempt to identify problems, the "specific" forms assume that the nature of the problem is known. These models thus focus on a specific accident cause or set of causes specified in advance. They are useful in either a performance evaluation or performance prediction context where the object of the search is known in advance. Their major function, as in the case of the general model forms, is to provide accident frequency counts for use in further analysis.

b. The second set of methods, Intervention Analysis, centers on the use of autoregressive models similar to the Box-Jenkins ARIMA models described in the preceding section. Known as Box-Tiao models, they are useful for evaluating the impact of a given program intervention when there exists a sufficient amount of time series data both before and after the intervention in question.

c. The third set of methods, termed Assessment of Impact Diagrams (AID), are a form of intervention analysis developed by Wyle Laboratories. Unlike the Box-Tiao models, they do not require precise time series observations; the data are simply dealt with as "before" and "after" aggregates.

2.3.1 Model Forms for Specific Analyses

These are models used to distinguish "treatment" accidents (accidents in which the given countermeasure was specifically involved) from "control" accidents (accidents in which the countermeasure was not involved).

These models are often used in a "quasi-experimental" context, to overcome certain barriers to the use of true experimental methods. For example, if one were searching for ways to reduce collisions where boats are struck in darkness or limited visibility, one might consider two possible remedies:

(1) Educating boaters to more widespread use of their lights (if one thought that functional lights were sometimes not used through poor judgment).

(2) Requiring through regulation more powerful boat lights (if one believed that lights were being used, but were insufficiently bright to achieve adequate warning).

Exploring these remedies through true experiments, such as those used in medicine, psychology, and other disciplines, presents certain problems. One might choose, for example, to establish dangerous situations for boats, some of whose operators had been specifically educated about the use of boat lights (treatment group) while others had not (control group). Similarly, for light brightness, one could establish dangerous situations for boats, some of which had enhanced lighting (treatment group) while others did not (control group). The relative occurrence of accidents in the treatment and control groups would offer evidence as to whether the control group circumstances were dangerous or not and whether the proposed remedy (treatment) was effective and how effective. The quality and size of the experiments would govern the credibility of the results and one's confidence in the numerical findings.

Where, however, such experiments are immoral or impractical (as in the cases cited), similar results can be achieved "quasi-experimentally." That is, accident data could be screened to identify incidents where a boat struck in limited visibility had functional lights which were not turned on. If a cluster of such accidents could be identified, some boaters similar to those in the accident cluster could be targeted for education about the use of boat lights. If the average occurrence of limited visibility strikings of such boats were to decline significantly in the group so educated, then one would have "quasi-experimental" evidence of an effective safety action.

A similar result might be achieved with no experimental action if one type of boater education course were known to stress use of lights and if boaters taking that course were found to be involved in the relevant accidents much less frequently than other boaters. Despite certain limitations inherent in this form of evidence (for example, does a boater become a safer operator because of the course or do those who take the course tend to be safer operators to begin with?), presumptive evidence of this nature, combined with other quasi-experimental results, can often present a compelling case in which each source of evidence serves to reinforce the others.

Despite the almost commonsense nature of these methods, they have often proved to be both difficult and expensive to formulate and apply. Data may not be available; different analysts may interpret data differently; and the appropriate exposure data may be lacking.

Section 1.2.3.2 of the Wyle report identifies three model forms suitable for use in specific analyses:

(a) The Casualty Analysis Gauge is suitable for use when only one particular accident/recovery path is of interest. Either that path is "met" or it is not. If met, the specific failure cause associated with that path is deemed to be present, and the case "counts" toward whatever study is being conducted.

(b) The Relative Occurrence Model is used when one is interested in several paths, all emanating from the same basic cause. If one of those paths is present, the case "counts."

(c) The Rating Model is used when one is interested in the relative frequency of several different causes, specified in advance. The output of this model is the relative frequency of each of those causes.

The Casualty Analysis Gauge would be used, for example, to study the relative incidence of accidents resulting from excessive powering at startup, the Relative Occurrence Model to study the incidence of powering problems in general, and the Rating Model to study the relative incidence of powering problems versus improper loading.

2.3.2 Intervention Analysis

Intervention analysis, as described in this section, makes use of models developed in the mid-1960's by Box and Tiao.[1] Since their inception, Box-Tiao models have been used in a variety of contexts to measure the impact of countermeasures ("interventions") for which there exist both pre- and post-intervention data. They have been used in the past to study, among other initiatives, the effects of a Connecticut crack-down on speeding, anti-pollution measures in Los Angeles, and a number of automotive alcohol control programs.

A useful feature of these models is that they take sampling uncertainty, based on the amount and variability of the data examined, into account. Thus, in addition to deriving measures of impact, Box-Tiao models permit one to place upper and lower confidence limits on the measure derived and to test the measure for statistical significance. For example, runs conducted by Wyle on 1969-1978 boating fatality data, based on an assumed intervention date of January 1974,[2] showed that there were an estimated 921 lives "saved" between 1974 and 1978 as a result of the intervention in question, with 90 and 95 percent confidence limits as follows:

	Lower Limit	Upper Limit	
90% confidence limits:	89	1,753	(i.e., 921 ± 832)
95% confidence limits:	-71	1,913	(i.e., 921 ± 992)

The width of these confidence bounds is a measure of the uncertainty inherent in the data. These numbers denote the following: if the process by which they were generated were to be repeated indefinitely, 90% of those repeated trials would yield an estimated impact of between 89 and 1,753 lives saved; 95% would yield an estimated impact of between 71 lives <u>lost</u> and 1,913 saved. The fact that the 90% lower limit is greater than zero, while the 95% limit is not, constitutes (in Wyle's words) results which are of "marginal significance".

---

1 Box, G.E.P. and Tiao, G.C., "A change in level of a non-stationary time series." Biometrika, 1965, 52, pp. 181-192.

2 The nature of the intervention is not clear from the text.

Another feature of the Box-Tiao models is that they can be run either with the original variables or with the data suitably transformed. When the runs just described were repeated with the data expressed logarithmically, Wyle derived an estimate of 1,289 lives saved, about 40% higher than that (921) obtained using the untransformed data.

The wide range between upper and lower confidence limits in the example given by Wyle is not surprising. Data instability is an ever-present threat in times series analysis and applies whether one uses Box-Tiao models or other, more conventional techniques.

Alternatives to the use of Box-Tiao models include calculating regression slopes before and after the intervention, followed by a statistical test to determine if the two regression coefficients differ significantly. Methods for performing these tests may be found in the SPSS and BMD references cited on page 21.

2.3.3 Assessment of Impact Diagrams

Another alternative to the use of Box-Tiao models is the Assessment of Impact Diagrams (AID) approach developed by Wyle. Unlike Box-Tiao, the AID approach deals with numerical aggregates (total number of fatalities, etc.) rather than explicit time series data. Essentially, AID says the following: if a given intervention was designed to impact upon a particular safety statistic X, one way of measuring that impact is to (a) find some other variable Y, related to X but which the intervention would not be expected to affect, then (b) compare the pre-and-post change in X to the pre-and-post change in Y. Thus, if Y (the control variable) declined by 10%, one would expect X (the treatment variable) to decline by at least that amount. The extent to which X declined beyond 10% is the estimated impact of the countermeasure in question.

EXAMPLE: In an attempt to reduce fatalities due to excessive powering, boats of a certain class built in 1973 or later are required to carry a maximum horsepower label. Reports on 103 pre-intervention fatalities show that 46 of those fatalities were related to powering and 57 were not. Of the 68 post-intervention fatalities for which data are available, half (34) were related to powering and half were not.

The following two-way contingency table, adapted from page 1-120 of the Wyle report, shows the situation:

TABLE A

	NUMBER OF FATALITIES	
	POWERING RELATED	NOT POWERING RELATED
PRE (Before 1973)	46	57
POST (1973 or later)	34	34

The principle behind AID is as follows: If labeling had had no effect, the pre- and post- figures in the left-hand column should bear the same ratio as the pre- and post- figures in the right-hand column; i.e., in the same way that fatalities which were not related to powering declined from 57 to 34, powering-related fatalities should have declined from 46 to 34/57ths of that amount, or 27.4. Since the decline in powering-related fatalities was less than expected -- i.e., the observed number of fatalities (34) was greater than the expected number of fatalities (27.4) -- labeling cannot be said to have had a positive impact.

In the preceding case, there was no need for sophisticated mathematics since the treatment group clearly failed to do as well as the control group. Suppose, however, that the number of powering-related fatalities had been less than expected, as shown in the following table:

TABLE B

	NUMBER OF FATALITIES	
	POWERING RELATED	NOT POWERING RELATED
PRE (Before 1973)	46	57
POST (1973 or later)	20	34

In this case, the treatment variable declined by a greater percentage than the control variable, implying that the labeling requirement may have had an impact. The question arises: is this greater-than-expected decline statistically significant or might it be due to chance?

To answer this question, Wyle defined a measure termed the effectiveness index e. The formulation given this variable in the Wyle report is more complex than required: essentially, e is simply the estimated impact derived above (i.e., the difference between the number of fatalities expected and the number observed) divided by the number of fatalities expected. In the case of Table A,

$$e = \frac{27.4 - 34}{27.4} = -0.24$$

while in the case of Table B,

$$e = \frac{27.4 - 20}{27.4} = 0.27$$

Positive values of e denote positive impact ("lives saved"); negative values denote negative impact ("lives lost"). The value calculated, however, is merely a sample estimate based on limited data; it is not necessarily statistically significant nor does it necessarily represent the "true" value of e. To address this uncertainty, Table 1-14 of the Wyle report proposes a test statistic for confirming or refuting the hypothesis that the true value of e is greater than some threshold value e_o which lies between zero and one. A simplified version of the test statistic proposed by Wyle is presented below:

$$h = \frac{\ln(1 - e_o) - \ln(1 - e)}{\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}$$

where ln denotes the natural logarithm, e is the calculated effectiveness index, e_o is the defined threshold value of interest, and a, b, c, and d are the four entries in the two-way contingency table.[1]

The test statistic h has several important properties. First, it is a standard normal variable, i.e., it is normally distributed with a mean of zero

[1] It makes no difference which value -- a, b, c, or d -- is assigned to which cell of the contingency table. The four cells are equivalent in this formulation.

and a standard deviation of one.[1] This means that any value of h greater than 1.96 is significant at the 95% level of significance.[2]

A second important feature of h is that it permits one to test the sample value of e not only against zero (i.e., for any impact whatsoever) but against any other value of interest to the analyst or decision-maker.

ILLUSTRATION:

On page 1-121 of the Wyle report, the following table is presented:

	NUMBER OF FATALITIES	
	POWERING RELATED	NOT POWERING RELATED
Boats which are not "safely" powered	20	10
Boats which are "safely" powered	26	47

(1) Calculate the effectiveness index e.

(2) Test the value of e for significance against a threshold value of 0.2.

SOLUTION:

(1) Since the purpose of safe powering is to reduce the incidence of powering-related fatalities, the variable of interest is the one in the lower left-hand corner. If safe powering were not really effective, the expected value of this variable would be (47/10) x 20, or 94; i.e., there would be 94 powering-related fatalities in boats which are safely powered. In truth, the observed value was 26. The effectiveness index is therefore

$$e = \frac{94 - 26}{94} = 0.72$$

---

1 Another way of stating this is that the variable ln (1-e) has a normal distribution with a mean of ln $(1-e_0)$ and a standard deviation equal to $\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$.

2 The reader is assumed to be familiar with tests of statistical hypotheses, particularly those involving normal variables. For those who are not, any standard statistical text will do.

SOLUTION (cont'd):

(2) The test statistic h is calculated as follows:

$$h = \frac{\ln(1 - .2) - \ln(1 - .72)}{\sqrt{\frac{1}{26} + \frac{1}{47} + \frac{1}{20} + \frac{1}{10}}} = \frac{1.0499}{0.4579} = 2.29$$

Since the calculated value of h is greater than 1.96, the analyst is justified in concluding that safe powering is characterized by an effectiveness index of at least 0.2.

Finally, the Wyle report deals with the issue of confidence intervals. Table 1-15 of the report defines the upper and lower confidence limits for e as follows:

$$\text{Upper limit:} \quad 1-(1-e)\exp\left[-z\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}\right]$$

$$\text{Lower limit:} \quad 1-(1-e)\exp\left[\ z\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}\right]$$

where z is the critical value of a standard normal variable corresponding to the desired level of confidence (i.e., z equals 1.96 for 95% confidence, 1.64 for 90% confidence, etc.).

In the illustration just presented, the 90% upper and lower confidence limits for e are seen to be

$$\text{Upper limit:} \quad 1 - .28 \exp[-1.64\ (.4579)] = .868$$

$$\text{Lower limit:} \quad 1 - .28 \exp[\ 1.64\ (.4579)] = .407$$

The fact that the lower limit is substantially greater than 0.2 confirms the significant findings noted at the top of this page.

Summary

Several sets of methods for evaluating boating safety performance are presented in this section. All are quasi-experimental in nature, i.e., they attempt to distinguish between events occurring to treatment and control groups (or to "pre" and "post" groups) without actually conducting experiments.

Two of the major sets of methods described in this section are of the pre/post-intervention variety. The first set, Intervention Analysis, makes use of mathematically elegant models which require substantial amounts of time series data. The second set, Assessment of Impact Diagrams (AID), is suitable where aggregate numbers exist for both the impact variable and a second variable which can be used as a control. Each of these methods presents some degree of uncertainty:

a. Intervention Analysis lacks a suitable control group. One cannot be certain the observed pre/post changes would not have taken place in any event.

b. AID <u>includes</u> a control group but the suitability of that group needs to be carefully examined. The assumption that the control variable would, in the absence of any impact, change at the same rate as the treatment variable may not always be valid.

3. MULTIATTRIBUTE UTILITY ANALYSIS

In pursuing the applications described in the preceding chapter, the analyst is often faced with a situation in which several variables -- not just one -- enter the decision process. Multiple program objectives may require simultaneous considerations: it may be necessary, for example, to trade off cost versus performance, or one aspect of performance versus another. Tradeoffs of this nature arise whether the analyst is involved in predicting performance (Section 2.2) or evaluating it (Section 2.3).

One approach, where there are only two variables involved, is to fix one variable (e.g., cost) and attempt to maximize the other. Where there are more than two variables, however, fixing all but one becomes an arbitrary task divorced from the decision maker's need for total optimization.

Multiattribute utility (MAU) analysis is a technique designed to handle the multi-objective situation. Chapter 3 explains the technique and outlines the steps involved in applying MAU methods to a typical boating safety problem. Readers interested in additional detail are referred to the bibliography at the end of this chapter.

3.1 Introduction

3.1.1 Background. Multiattribute utility analysis is useful for any decision in which multiple factors, or attributes, are important, no decision is clearly best on all factors, and some factors are difficult to quantify. It is best applied to situations in which there is a well-defined set of alternatives that differ on the attributes. Typically the attributes must be traded-off among each other such as cost vs. benefit, reward vs. risk, long vs. short term, and effectiveness vs. political considerations. Such tradeoff issues are often a matter of personal preference of the decision maker, and as such, subjective judgments become a critical part of the analysis.

This chapter is intended to be a user's manual for MAU applications. The thrust is tutorial in nature, and is oriented primarily for the analyst who is relatively inexperienced regarding such methods. All aspects of an MAU will be introduced in the context of an example problem that is typical of those facing U.S. Coast Guard analysts. When appropriate, more complex aspects of an MAU will be referenced for use by the more sophisticated user.

3.1.2 Example problem to illustrate the methodology. Coast Guard boating safety statistics show that in 1982, 1178 people lost their lives in recreational boating accidents. Almost 90% of these deaths were the result of drowning. To reduce the number of such deaths, the Coast Guard has established standards and requirements for personal flotation devices (PFDs). In order to demonstrate the MAU methodology, the following hypothetical evaluation problem will be examined. In this example, Coast Guard analysts have been tasked with examining the way that boaters evaluate several alternative PFDs with regard to major factors such as effectiveness and cost. There is insufficient time and dollars available for a detailed, in-depth test program; however, there is some data and many experts available that are knowledgable on the PFDs. The specific goal of the analysis is to provide a summary evaluation of the alternative PFDs and to identify the major issues that affect boaters' choices.

(Note that the scope of the sample problem is scaled down to illustrate the MAU technique. The following analysis is not intended to be an actual evaluation of PFDs.)

3.2 The Methodology

3.2.1 Overview of MAU. The key stages in an MAU approach, as they relate to the example problem of Section 3.1.2, are as follows:

- Problem Definition (how do boaters select PFDs?);
- Identification of what is to be evaluated (alternatives or options, i.e., what are the PFDs from among which boaters can choose?);
- Definition of the components, or attributes of value (what is important, e.g., cost, effectiveness, durability of PFDs?);
- Evaluation, or "scoring" based on the attributes (how is each PFD rated on each attribute?);
- Prioritization of the attributes of value (e.g., is cost more important than efficiency in selecting a PFD?);

- Comparison of alternatives being evaluated (which PFD scores highest on all factors combined?);
- Sensitivity analysis on assumptions and judgments (what if priorities change?).

3.2.2 Problem definition. As indicated above, the most frequently applied use of MAU is for evaluation of a well-defined set of alternatives. The nature of the problem is such that the analyst must compare alternatives and select from among them based upon evaluation scores. Not only would the analyst like to know how alternatives compared when all factors are considered, but he should also be able to identify readily the contribution of the factors to the overall evaluation.

In the example problem, the problem definition could be stated as "determine the factors that lead recreational boaters to select one PFD over another, and evaluate a specified set of PFDs based upon such factors."

3.2.3 Identification of the alternatives. In many cases, the alternatives to be evaluated are few and well-defined. In other cases, it becomes necessary to pare down the potential set of alternatives before detailed evaluation. This is often accomplished by a technique known as elimination by aspect. For example, in buying a car, few buyers fully evaluate all models. Rather, they eliminate many alternatives outright by specifying certain required and/or unacceptable aspects of the automobiles. This might include specifying a price range, a style (e.g., 2 door, convertible), a manufacturer, or even specific features (e.g., must have automatic transmission).

In generating options for cases where alternatives are not well defined, it is sometimes useful to focus on one characteristic that plays a major role in the decision. This characteristic is used to generate different alternatives as it runs through its range of potential values. For example, alternatives can be characterized from least risky to most risky, cheapest to most expensive, easiest to implement to hardest to implement, etc.

For our example problem, we will look at a subset of potential PFDs that are to be evaluated. These are:

(1) Inflatables (I):
(2) Hybrid inflatables (HI);
(3) Inherent buoyancy devices (IB);
(4) Styled inherent buoyancy (SIB);
(5) Ski belt (SB)
(6) Seat cushions (SC).

This illustration is intended to demonstrate the methodology. It is not intended to be a complete and valid evaluation of PFDs.

3.2.4 Identifying attributes of value. In determining the attributes, or factors, on which the alternatives will be judged, it is desirable that the set of attributes have the following characteristics:

- Be comprehensive enough to account for most of what is important in evaluating the options;
- Be able to highlight the differences among options;
- Reflect separate, nonoverlapping features to avoid double counting.

While it is desirable to satisfy the last characteristic, it is by no means required. It is possible to define evaluation factors that are dependent upon each other and interact in complex ways. However, most of the value of an MAU model can usually be obtained by using a simpler form in which each factor is independent of all other factors. If it is clear that two factors are not independent, but both are interacting, it is usually possible to define a single factor that incorporates the critical aspects of the dependent factors. For the purposes of this manual, all models will be structured to have independent factors.

It is relatively easy to define the attributes in a hierarchical fashion such that at the top of the hierarchy are broad, general attributes which get subdivided into more specific sub-attributes. Usually, the highest level attributes are too broad to be useful in scoring alternatives; thus, the rule-of-thumb for subdividing is to develop attributes at the lowest level of the hierarchy that can be measured readily. A simple hierarchical evaluation structure for the PFD evaluation example is shown in Figure 3.1.

There are four major factors: effectiveness of the device, the usability of the PFDs by boaters, the durability of the PFDs themselves, and the costs of the PFDs. If the analyst truly wants to discriminate among the options, these factors are too broad. As a result, they were subdivided as indicated:

Effectiveness

- HEAD-UP - the ability to keep the head out of water for an unconscious boater;
- ROUGH SEAS - the ability to keep a conscious boater afloat in rough seas;
- BUOYANCY - the rated weight capacity of the device;
- FAILURE RATE - indicates potential for not doing the job for which it was intended (e.g., inflatable fails to inflate).

Usability

- IMAGE - measures how boaters perceive their own image of using the device;
- ACCESSIBILITY - the ability to store the device and get to it readily when needed;
- COMFORT
  - WEARABILITY - measures how comfortable the device is to wear
  - INTERFERENCE - indicates how much the device interferes with other activities such as fishing, sunbathing, etc.

Durability

- SHELF LIFE - how long can the PFD remain on the shelf unused before it begins to deteriorate?
- EXPOSURE LIFE - how long can the PFD be used and exposed to boating conditions before it begins to deteriorate?
- ROUTINE USE - how susceptible is the device to damage caused by routine use (e.g., straps tearing, punctures, etc.);
- REUSABILITY - once used, is the ability to reuse the PFD impaired?

Cost

- INITIAL PURCHASE COST;
- O&M COSTS - costs to operate, maintain and replace the PFD (over a 10-year period).

Each factor in the "tree" structure is referred to as a branch, and the places where branches meet are referred to as nodes. Note that there is no requirement that the number of levels of subdivision be the same throughout the structure, nor do all nodes necessarily have the same number of branches. The last level of subdivision results in branches that are called bottom-level attributes, or terminal branches. The terms "factors," "attributes," "branches," and "criteria," are generally interchangeable, and all are commonly found in the literature on MAU.

There is virtually no limit to the number of levels of the hierarchy or the number of bottom-level attributes that can be developed as an MAU structure. However, as a general guideline, five levels are usually more than adequate. If in structuring the problem, the analyst has more than five levels, he should give serious consideration to regrouping attributes. A normal tendency for the beginner is to attempt to develop a "tree" with minute levels of detail to ensure that nothing is left out. The analyst must remember that the primary purpose of MAU is to differentiate among alternatives. Attributes that provide no contribution towards differentiation should be considered for elimination.

There are many techniques that can be used to develop the tree structure, but two are most prevalent -- top-down structuring and bottom-up structuring.

In top-down structuring, the analyst first describes the highest level attributes and then attempts to determine logical subdivisions. The analyst proceeds from general to specific until a level is reached that provides a reasonable measure of value. It is perhaps the easiest technique to use and as such, is popular with less experienced users.

The bottom-up approach is more difficult to employ, but often results in a more discriminatory structure. The idea is to generate the lowest-level attributes by directly identifying measurable factors, and then logically grouping the factors into clusters that go from specific to general. One of the best ways to apply this approach is to begin by listing the advantages and disadvantages of each alternative. The lists for all alternatives are then combined into a single list of advantages and disadvantages. This list is then used to define the attributes and to group them into logical clusters. The bottom-up approach is less likely to miss an attribute inadvertently, but is more time-consuming and requires more experience than the top-down approach.

With either approach, the following issues are critical to remember:

(1) At any node of attributes, the branches should be independent; all nodes should be independent of each other.

(2) Minute detail is usually unnecessary; focus on attributes that discriminate among options. Typically, one-third of the bottom-level attributes account for 80-90% of the overall evaluation.

(3) When properly used, differences in numbers of levels and differences in number of branches at nodes will not affect the results.

(4) In determining if an attribute can be easily measured, it is not necessary that the attribute have an obvious objective measure that can easily be quantified. Measures that must be evaluated subjectively using expert judgments are equally valid and should be used. (Benefit assessment techniques are discussed in a later section.)

3.2.5 Evaluation of alternatives on attributes. In order to use the MAU structure for evaluation purposes, it is necessary to develop a measurement scale for each bottom level attribute. Such measurement scales should be developed using natural standard units whenever possible (e.g., dollars, years), but it is often necessary to use more subjective, relative scales. Two major procedures for scoring will be described here -- relative scoring procedures and absolute scoring procedures.

Relative scoring is perhaps the easiest technique to use, but requires that the alternatives are clearly specified. The common measurement of value that is used is relative utility, and it can be measured on a scale from 0 to 100. These endpoints are somewhat arbitrary in that virtually any endpoints can be used; however, once fixed, the endpoints serve as a reference point for other assessments. In relative scoring, for each factor, the alternative that is "best" on the factor is assigned a score of 100, while the "worst" alternative is assigned a score of 0. The range of such a scale thus measures the difference between options -- a score of 100 can be thought of as 100% of the potential improvement on a factor over and above the baseline worst case which scored 0. Note that a score of 0 does not imply that the alternative has no value. Rather, it indicates that the alternative is the baseline for comparison. All other alternatives are scored on the 0 to 100 scale relative to how they compare with the endpoints. A score of 50 on the above defined 0 to 100 utility scale means that the satisfaction level, or utility of the alternative is midway between the best and worst. Note that a score of 100 on one attribute cannot be directly compared with a score of 100 on another attribute since they may not be equally important. In order to compare such attributes, a weighting procedure must be applied as described in a later section.

An an example of relative scoring, consider the bottom-level attribute of DURABILITY -- SHELF LIFE. This scale would have the obvious measure of years. The utility scale will thus serve to assign a score to the number of years associated with an alternative. Suppose the following SHELF LIFE data are available on the options: (Note: all data used in this section are hypothetical.)

Option	SHELF LIFE (years)
I	3
HI	5
IB	10
SIB	10
SB	6
SC	8

It is reasonable to assume that the value of SHELF LIFE is linear with time. The best options are IB and SIB, therefore both score 100. The shortest SHELF LIFE is option I which is given a score of zero. The other options are scored relative to these endpoints. SB with a SHELF LIFE of 6 years should be 3/7 of the way from 0 to 100, or a score of 43. This is calculated as:

$$\frac{\text{Distance from endpoint}}{\text{Range}} = \frac{6-3}{10-3} = \frac{3}{7}$$

Similarly, SC would score 71:

$$\frac{8-3}{10-3} = \frac{5}{7} \longrightarrow 71.$$

All values can be shown on a utility curve as in Figure 3.2.

A utility curve reflects the preferences of the "decision maker," and, as such, is a very personal representation of value. These preferences can be obtained using interviews, surveys, or other assessment techniques. As an example of the personal nature of such values, it is logical that a very wealthy boater would have a utility curve on costs of PFDs that is very different from a boater with much less to spend. In fact, each individual boater could have a unique set of curves. For the PFD example, we will assume that the analyst is evaluating PFDs using the preferences of the "average" boater as obtained from survey data. As part of a sensitivity analysis, results for other boater profiles could be examined.

EXAMPLE FOR ILLUSTRATION ONLY

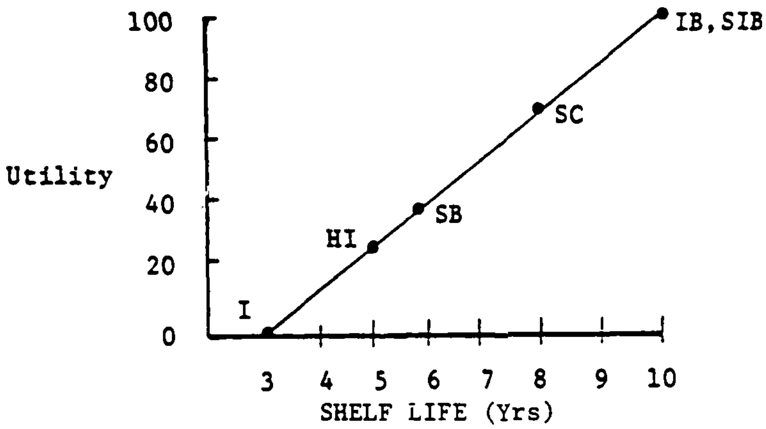


Figure 3.2: Utility Curve for SHELF LIFE

EXAMPLE FOR ILLUSTRATION ONLY

There is no requirement that a utility curve be linear. For example, on the factor BUOYANCY, it can be argued that the initial improvements in BUOYANCY over baseline have the most incremental value. Suppose BUOYANCY measures are as follows:

Option	BUOYANCY (lbs)
I	25
HI	25
IB	25
SIB	17
SB	5
SC	9

If the argument can be made that the eight pound difference in going from SC to SIB is more important than the eight pound difference in going from SIB to IB, the utility curve might look like the following:

EXAMPLE FOR ILLUSTRATION ONLY

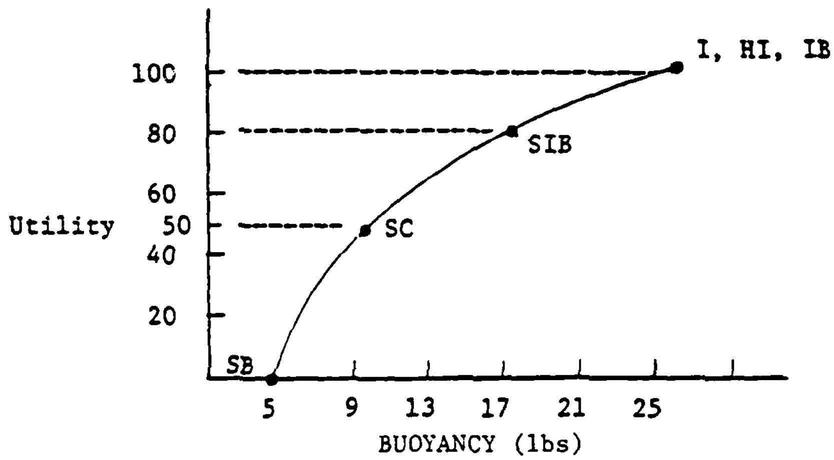


Figure 3.3: Utility Curve for BUOYANCY

EXAMPLE FOR ILLUSTRATION ONLY

The utility curve can, in fact, take on many different shapes. In some cases, utility increases slightly until a threshold is reached and then it rises dramatically (e.g., HEAD-UP FLOTATION). In other cases, utility is "all or nothing;" that is, no value is perceived until a certain point is reached, then all value is obtained (e.g., can the PFD be reused?). It is also theoretically possible for utility to rise up to a point and then drop off (e.g., if it were the case that too much buoyancy keeps the body out of water and causes exposure problems). These situations could lead to the following types of curves:

EXAMPLE FOR ILLUSTRATION ONLY

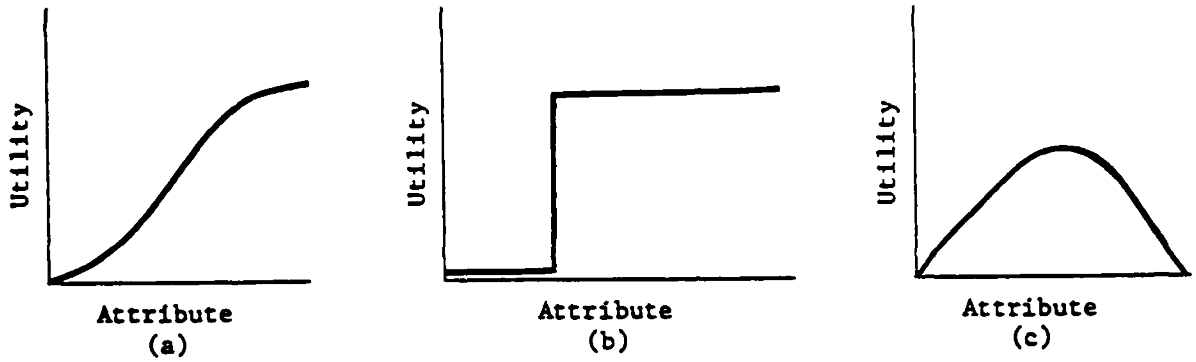


Figure 3.4a, b, c: Possible Shapes for Utility Curves

EXAMPLE FOR ILLUSTRATION ONLY

There is also no requirement that utility curves be continuous. Often, the attribute can be measured in discrete terms, even though there is a continuous range for the measure. For example, the factor INTERFERENCE could be measured as follows:

EXAMPLE FOR ILLUSTRATION ONLY

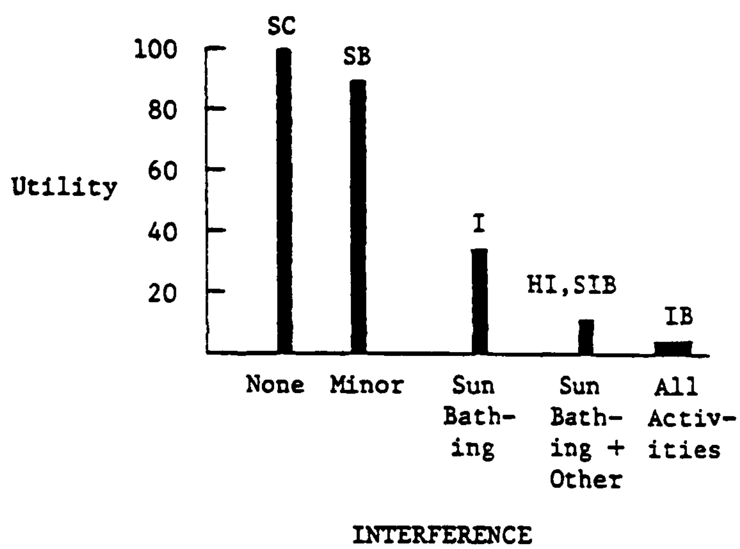


Figure 3.5: Discrete Utility Curve for INTERFERENCE

EXAMPLE FOR ILLUSTRATION ONLY

In the absolute scoring system, the scales are used in a similar fashion to the relative scoring system, but alternatives can fall anywhere on the scale. The utility scales again can take a variety of forms and shapes, and a weighting system is required to compare one scale with another.

The absolute scoring system usually provides results that are easily communicated, and provides a better reference frame of value than relative scales. It also allows adding or deleting of options without requiring revision to the utility curves. On the other hand, an absolute system is more difficult to use, requires more substantive expertise on the nature of the problem, and is far more time consuming to develop. The absolute scoring system is best used in situations where alternatives are not well defined. In fact, it serves as a useful tool in developing feasible alternatives, and can be used to recommend where appropriate data are readily available.

In the example problem on PFDs, assume that the following measures of value have been defined for the bottom-level attributes using the relative scoring procedure:

Attribute	Measure
HEAD-UP	Relative scale - best to worst
ROUGH-SEAS	Relative scale - best to worst
BUOYANCY	Pounds of buoyant force
FAILURE RATE	Failures/hrs use
IMAGE	Relative scale - best to worst
ACCESSIBILITY	Relative scale - best to worst
WEARABILITY	Relative scale - best to worst
INTERFERENCE	Relative scale - best to worst
SHELF LIFE	Years
EXPOSURE LIFE	Years
ROUTINE USE	Relative scale - best to worst
REUSABILITY	Relative scale - best to worst
INITIAL COST	Dollars
O&M COST	Dollars (discounted life cycle)

Table 3.1: Measures for the Bottom-Level Attributes

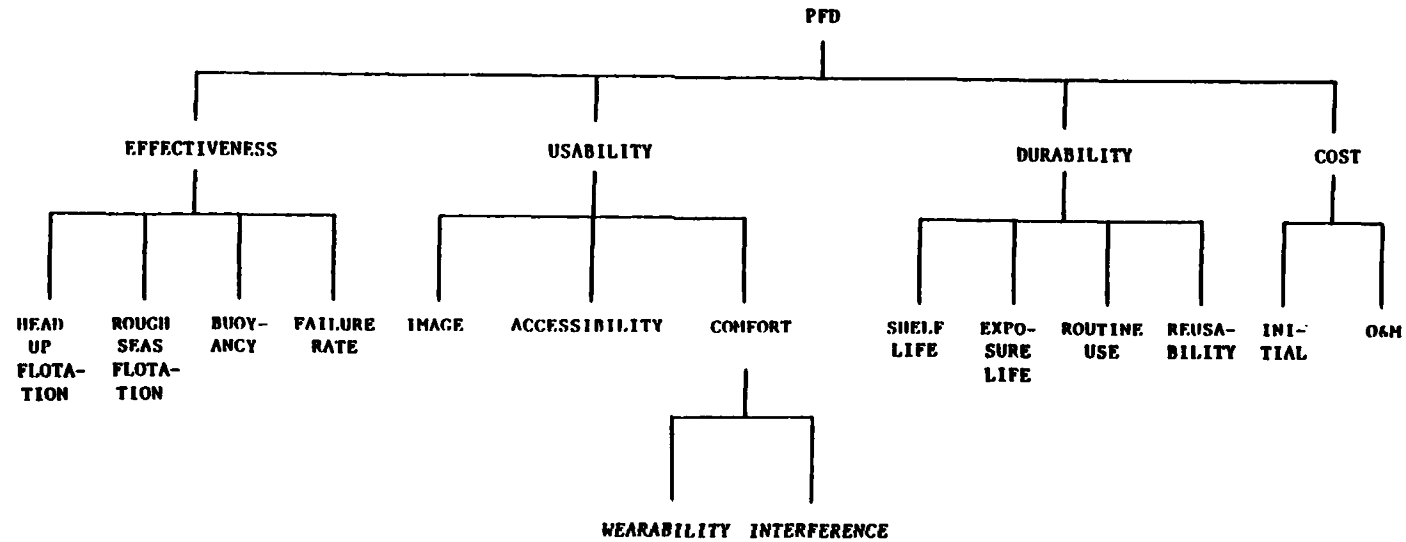
Again, using relative scoring procedures, scores can be developed for the alternatives on each attribute. For example, in evaluating INITIAL COST, seat cushions (SC) are cheapest at $10, thus they score 100. Stylized inherent buoyancy (SIB) is most expensive at $50 and scores 0. Inherent buoyancy (IB), at $40, is one-fourth of the cost range from SIB to SC, thus IB scores 25. Similarly, with inflatables (I) at $35, hybrid inflatables (HI) at $45, and ski belts (SB) at $12, their respective scores would be 38, 12, and 95. Of course, this assumes a linear relationship between dollars and utility. There is no requirement for such an assumption; however, it seems appropriate in this case.

Scores for all alternatives on all bottom-level attributes are displayed using the previous tree structure as shown in Figure 3.6. Recall that these scores are hypothetical and do not represent Coast Guard judgments. Rather, they have been developed by the authors to demonstrate key points of the methodology.

3.2.6 Prioritization of the attributes (weighting). In the scoring systems described above, an evaluation scale from 0 to 100 was developed for each factor. However, each scale is defined independently of all others, and the resulting scores are not directly comparable. In reality, some attributes carry more importance in the evaluation than others, and a measure of the priority, or relative importance, of each factor is necessary. This is accomplished through a weighting system. As with the scoring system, weighting judgments are personal, and different decision makers could have different sets of weights. As indicated earlier, in this example, we will assess weights for the "average" boater and treat other possible weights through sensitivity analysis.

The most common perception of a weight is that it answers the question "How important is attribute A relative to attribute B?" Unfortunately, such a measure is often inadequate in providing good discrimination among options. A more pertinent question to ask is "How important is the difference along the range in values for attribute A versus the difference for attribute B?" The subtle difference in wording of these two questions is extremely important. The latter question includes both the importance of the attribute as well as the "swing" in the range of values on the attributes. As an example of this distinction, assume you are evaluating three new job opportunities on three attributes -- job satisfaction, location, and salary. Without considering the specific alternatives, your judgment might be that salary is most important, then location, then job satisfaction. However, in looking at the three new jobs, you discover that

EXAMPLE FOR ILLUSTRATION ONLY



I	50	50	100	0	50	100	40	30	0	0	0	0	38	12
HI	50	60	100	10	20	80	10	15	29	17	20	10	12	0
IB	100	100	100	70	0	0	0	0	100	100	100	100	25	100
SIB	5	25	80	100	75	20	50	15	100	67	90	100	0	81
SB	0	0	0	95	90	90	80	90	43	33	50	100	95	84
SC	0	0	50	85	100	80	100	100	71	50	60	100	100	94

Figure 3.6: PFD Evaluation Structure with Scores

EXAMPLE FOR ILLUSTRATION ONLY

all pay the same salary. Intuitively, it is clear that salary will not be the most important factor in the decision since there is no difference among the options. A realistic set of weights would measure the importance of the differences in attributes among options rather than the importance of the attributes alone. For the remainder of this discussion, whenever the importance of an attribute is mentioned, we are referring to the "swing" importance.

The same notion of weighting holds for both relative and absolute scoring scales. In the former, the differences between best and worst alternatives are the basis for weights, while in the latter, it is the difference between the selected endpoints of each scale. In either case, if the true tradeoffs are to be captured, the "swing" weight notion is essential.

As in the case of structuring, weighting can be accomplished top-down or bottom-up. Top-down weighting is easier and is typically used by less experienced analysts. In the top-down approach, the analyst begins at the highest level node in the hierarchy, and assesses the relative differences among attributes. For the PFD example, the questions might be "How do differences among options in EFFECTIVENESS compare in importance with differences in USABILITY, DURABILITY, and COST?," or, "Is it more important to get improved capability over baseline in USABILITY versus EFFECTIVENESS, DURABILITY, and COST?" One of the more common approaches is to assign a weight of 100 to the most important swing. Other weights are then assigned using ratio judgments -- that is, if the swing on an attribute is judged to be twice as important as the swing on another attribute, the former would carry twice the weight of the latter. Due to the independence assumptions discussed earlier, weights can be compared in an additive sense. If attribute A is weighted at 100, attribute B at 75, and attribute C at 50, this implies that improving on B and C together (added weights equal 125) is more important than improving A (score of 100). This tends to serve as a good calibration check on the weights. The typical tendency for the inexperienced analyst is to group weights closely together, e.g. 100, 95, 90. When challenged by additive checks, such closely grouped scores normally are inconsistent with expressed judgments. Thus, by using additive comparisons as consistency checks, the weights get spread more widely to reflect the true relative importances.

Assume that differences in EFFECTIVENESS are determined to be most important, differences in COST are 3/4 as important as in EFFECTIVENESS and as important as USABILITY and DURABILITY combined, and differences in USABILITY and DURABILITY are equally important. Using the above procedures, the following swing weights were inferred from these judgments for the top level attributes in the PFD example: (Note: These weights do not represent Coast Guard assessments. They are hypothetical values developed by the authors.)

Attribute	Assigned Weight
EFFECTIVENESS	100
USABILITY	37.5
DURABILITY	37.5
COST	75

For convenience, the weights can be normalized to sum to 100 by adding the assigned weights and dividing each by the sum as shown below:

Attribute	(Sums to 100)
EFFECTIVENESS	40
USABILITY	15
DURABILITY	15
COST	30

Looking at USABILITY, these 15 points of weight must be spread among the subfactors that make up USABILITY. There are three branches making up USABILITY -- IMAGE, ACCESSIBILITY, and COMFORT. Rather than trying to allocate the 15 points, it is easier to assign weights to these subfactors using the same approach described above. The most important swing is given 100, ratio judgments are made for other swings, and the resulting numbers are normalized to sum to 100. Thus, at every node in the structure, the "local" weights will sum to 100. Assume that the USABILITY subfactor weights were judged as follows:

Attribute	Assessed Weight	Normalized Weight (%)
IMAGE	75	37.5
ACCESSIBILITY	25	12.5
COMFORT	100	50

This process would next be applied to the attribute COMFORT since it has subattributes of WEARABILITY and INTERFERENCE. Assigning weights would continue for all nodes in the structure.

The cumulative weight, or CUMWT, of an attribute is the product of all normalized weights along the branches leading to the attribute in question. To illustrate, the evaluation structure of Figure 3.6 is repeated in Figure 3.7 with hypothetical local weights shown for each attribute, as well as the CUMWT. To calculate the CUMWT for ROUGH SEAS FLOTATION, multiply the local weights along the path to ROUGH SEAS, or 40% (for EFFECTIVENESS) times 10% (for ROUGH SEAS) which equals 4% (.04). Note that CUMWTS for all bottom-level attributes sum to 100%. The interpretation of the .14 CUMWT for HEAD-UP FLOTATION is that in terms of the entire evaluation, the importance of differences among alternatives on that attribute accounts for 14% of the entire decision. As a further calibration check on the assigned weights, it is useful to list all bottom-level factors in order of decreasing CUMWT and to observe how much of the total weight is accounted for by an attribute and those preceding it. This provides a basis for discussion and revision of the assessed weights. For the PFD example, this list would appear as follows:

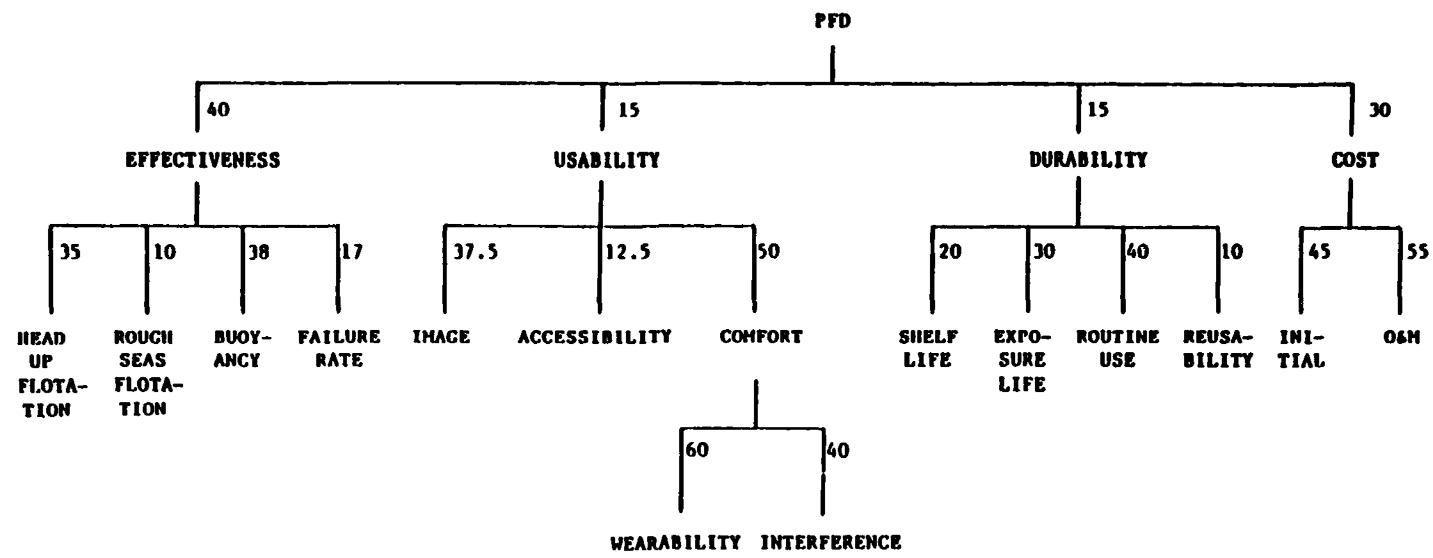
EXAMPLE FOR ILLUSTRATION ONLY

Bottom-Level Attribute	CUMWT	Sum of CUMWTs
O&M COST	.165	.17
BUOYANCY	.15	.32
HEAD-UP FLOTATION	.14	.46
INITIAL COST	.135	.60
FAILURE RATE	.07	.67
ROUTINE USE	.06	.73
IMAGE	.056	.78
WEARABILITY	.045	.83
EXPOSURE LIFE	.045	.87
ROUGH SEAS FLOTATION	.04	.91
SHELF LIFE	.03	.94
INTERFERENCE	.03	.97
ACCESSIBILITY	.019	.99
REUSABILITY	.015	1.00

Table 3.2: Attributes Prioritized by CUMWT

EXAMPLE FOR ILLUSTRATION ONLY

EXAMPLE FOR ILLUSTRATION ONLY



CUMWT	.14	.04	.15	.07	.06	.02	.04	.03	.03	.04	.06	.01	.14	.17
I	50	50	100	0	50	100	40	30	0	0	0	0	38	12
HI	50	60	100	10	20	80	10	15	29	17	20	10	12	0
IB	100	100	100	70	0	0	0	0	100	100	100	100	25	100
SIB	5	25	80	100	75	20	50	15	100	67	90	100	0	81
SB	0	0	0	95	90	90	80	90	43	33	50	100	95	84
SC	0	0	50	85	100	80	100	100	71	50	60	100	100	94

Figure 3.7: PFD Evaluation Structure with Scores
Top-Down Weighting

EXAMPLE FOR ILLUSTRATION ONLY

Note that the top 4 of the 13 attributes account for 60% of the total weight.

As indicated earlier, top-down weighting can be done quickly and with little computational difficulty. Its major disadvantage is that in making tradeoffs at the highest levels, it is difficult to conceptualize all of the things that are included in comparing differences among attributes. In essence, the analyst, in working with the expert must aggregate these factors implicitly which is a non-trivial task. Using top-down weighting, it often happens that one "slips" back to the <u>absolute</u> importance measure of a factor rather than the more desirable <u>swing</u> importance.

A more complex alternative for assessing importance weights is the bottom-up approach. The analyst begins at the lowest-level attributes and works his way upwards by directly comparing lower-level attributes in one part of the structure with attributes in another part.

For example, in the PFD problem, the analyst could start with the EFFECTIVENESS node and elicit weights for HEAD-UP FLOTATION, ROUGH SEAS FLOTATION, BUOYANCY, and FAILURE RATE as before. Assume the weights are as follows:

Attribute	Assessed Weight
HEAD-UP	90
ROUGH SEAS	25
BUOYANCY	100
FAILURE RATE	45

The analyst next compares an attribute from another part of the structure with one of the previously assigned weights. For example, assume BUOYANCY improvements (weighted as 100) are considered to be five times as important as INTERFERENCE improvements. The weight on INTERFERENCE would be one fifth that of BUOYANCY, or would be 20. Since in this approach, the attributes are all being evaluated on a common scale, the weights as initially assessed <u>can</u> be directly compared. Similarly, suppose ACCESSIBILITY swings are judged to be half as important as ROUGH SEAS swings, and HEAD-UP FLOTATION is three times as important as WEARABILITY. This could imply a weight of 12.5 for ACCESSIBILITY and 30 for WEARABILITY. A "subtree" indicating assessments made thus far is shown in Figure 3.8.

The power of bottom-up weighting becomes evident as we move up the hierarchy. Since all bottom-level attribute weights are linked through measurement on a common scale, it is only necessary to add together the weights of the branches at each node to get the node weight. Thus, the weight of EFFECTIVENESS is the sum of the weights of its components, or 90 + 25 + 100 + 45 = 260. Similarly, COMFORT weight is the weight of WEARABILITY (30) plus INTERFERENCE (20), or 50. If the weight of IMAGE is assessed at three times that of ACCESSIBILITY (12.5), the weight of USABILITY would be 37.5 (IMAGE) + 12.5 (ACCESSIBILITY) + 50 (COMFORT) or 100. Similarly, we might judge INITIAL COSTS to be slightly less important than HEAD-UP FLOTATION (90) so we could assign a weight of 85. If differences in O&M COSTS are judged to be slightly more important than improvements in BUOYANCY (100), we might assign a weight of 105. Note that in making such judgments, the process is iterative and multiple comparisons are made as consistency checks. In an oversimplified analogy, this can be compared to the process that an eye doctor follows in determining a prescription. Rather than asking "What is your vision?" (i.e., what is the weight of an attribute?), he asks "Do you prefer lens A or lens B?" (i.e., is attribute A more important than B?). He then continues by making additional comparisons.

EXAMPLE FOR ILLUSTRATION ONLY

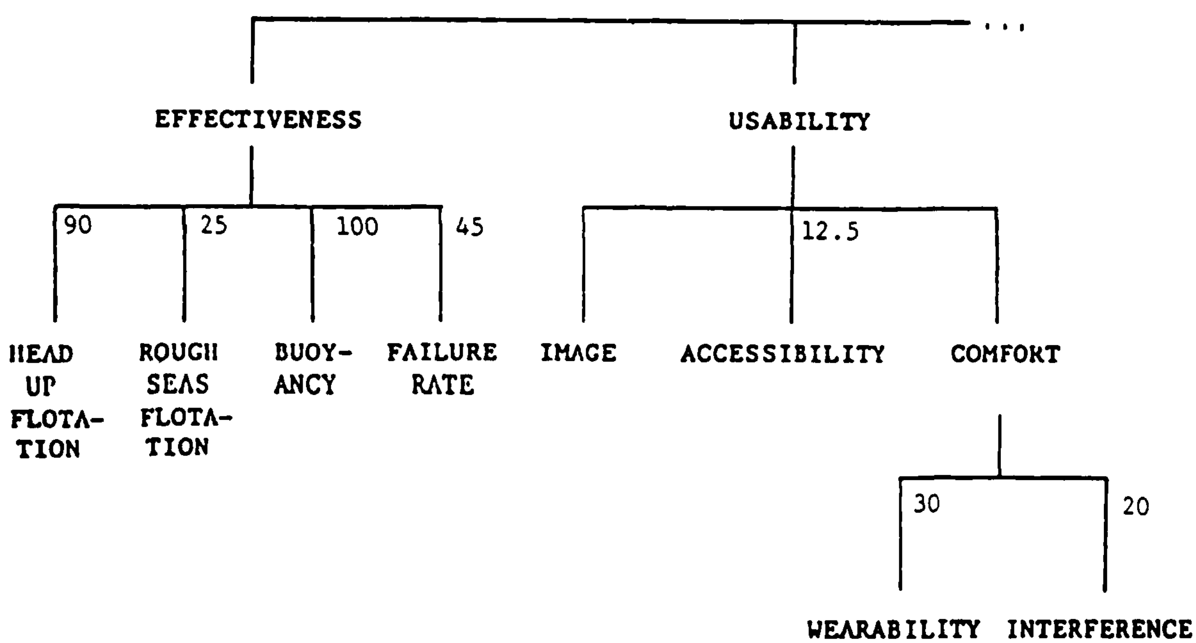


Figure 3.8: Subtree Illustrating Bottom-Up Weighting

EXAMPLE FOR ILLUSTRATION ONLY

Using a similar process for the rest of the structure, assume that weights have been assessed as shown in Figure 3.9. All weights as shown are directly comparable. Next, at each node, we can normalize weights to sum to 100 and can calculate CUMWTS as before. (Note that CUMWTS can also be calculated by adding all bottom-level assessed weights and dividing each weight by the sum.) The results are shown in Table 3.3. Note that for the purposes of this example, the results of both top-down and bottom-up weighting procedures are identical. In reality, it would be highly unlikely that this would occur. Perceptions change, many aggregate judgments are made implicitly, and there are few perfect judgment assessors in the world. However, results should be consistent. The key to this is the iterative process, the challenging of the final weights, and revision based on logical argument.

EXAMPLE FOR ILLUSTRATION ONLY

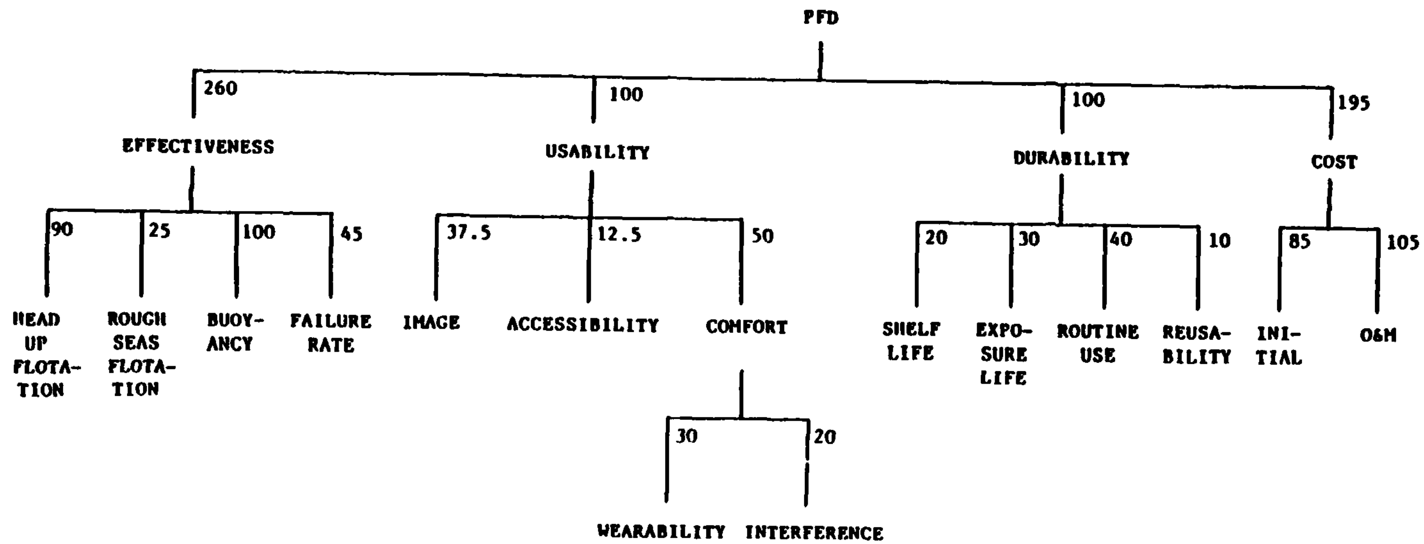
Attribute	CUMWT
HEAD-UP FLOTATION	.14
ROUGH SEA FLOTATION	.04
BUOYANCY	.15
FAILURE RATE	.07
IMAGE	.06
ACCESSIBILITY	.02
WEARABILITY	.04
INTERFERENCE	.03
SHELF LIFE	.03
EXPOSURE LIFE	.04
ROUTINE USE	.06
REUSABILITY	.01
INITIAL COST	.14
O&M COST	.17

Table 3.3: CUMWTS from Bottom-Up Weighting

EXAMPLE FOR ILLUSTRATION ONLY

In practice, it is likely that the analyst will find that a combination of both bottom-up and top-down weighting will work quite well. By using both techniques at different points in the weighting process, greater consistency can be achieved.

EXAMPLE FOR ILLUSTRATION ONLY



I	50	50	100	0	50	100	40	30	0	0	0	0	38	12
HI	50	60	100	10	20	80	10	15	29	17	20	10	12	0
IB	100	100	100	70	0	0	0	0	100	100	100	100	25	100
SIB	5	25	80	100	75	20	50	15	100	67	90	100	0	81
SB	0	0	0	95	90	90	80	90	43	33	50	100	95	84
SC	0	0	50	85	100	80	100	100	71	50	60	100	100	94

Figure 3.9: PFD Evaluation Structure with Scores
Bottom-Up Weighting (Unnormalized)

EXAMPLE FOR ILLUSTRATION ONLY

3.2.7 Comparison of the alternatives. After all alternatives have been scored on the attributes, and weights have been assigned to the attributes, the analyst must determine an overall measure of value for each alternative. Since the MAU described in this manual uses independent attributes, the overall score will be an additive combination of scores and weights as will be described below. In more complicated structures, where attributes do interact, a multiplicative model is more appropriate. Multiplicative models are discussed in Section 3.3.1.

It is essential to recognize that the numerical results of the evaluation process are not the ultimate goal of the model. Rather, the scores and weights are merely a reflection of the judgments used as inputs. The numerical output should serve the analyst as a catalyst for discussion and revision of the model. A perfectly acceptable (and often desirable) outcome of the MAU model is a result that is not intuitively appealing. The beauty of the MAU model is the ease with which such disagreements can be traced to specific rationale, and revised if appropriate. As such, the computational algorithm should be presented in a form that allows such traceability.

For ease in referring to various parts of the MAU model, each node and branch in the evaluation structure can be numbered in a hierarchical fashion as shown in Figure 3.10.

Each node also can be represented in matrix form as shown in Table 3.4 for node 1.1, EFFECTIVENESS:

EXAMPLE FOR ILLUSTRATION ONLY

NODE 1.1, EFFECTIVENESS		OPTION SCORES						
BRANCHES:	LOCAL WGT	I	HI	IB	SIB	SB	SC	CUMWT
1.1.1 HEAD-UP FLOTATION *	.35	50	50	100	5	0	0	.14
1.1.2 ROUGH SEAS FLOTATION*	.10	50	60	100	25	0	0	.04
1.1.3 BUOYANCY *	.38	100	100	100	80	0	50	.15
1.1.4 FAILURE RATE *	.17	0	10	70	100	95	85	.07
TOTAL	1.00	61	63	95	52	16	33	.40

Table 3.4: Matrix for Node 1.1

EXAMPLE FOR ILLUSTRATION ONLY

EXAMPLE FOR ILLUSTRATION ONLY

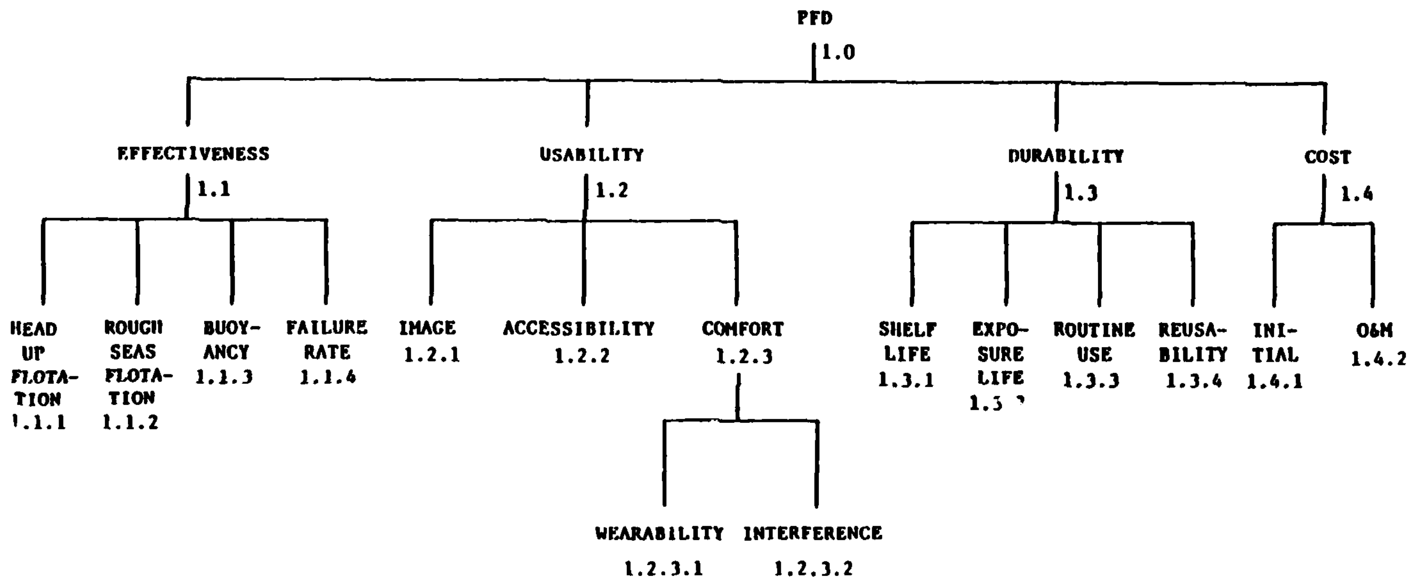


Figure 3.10: PFD Evaluation Structure with Nodes and Branches Numbered

EXAMPLE FOR ILLUSTRATION ONLY

The column labeled node weight would reflect the "local" weight of each branch at the node and will sum to 100%. The next six columns reflect the scores for each option as assessed in Section 3.2.5. The rightmost column reflects the CUMWT of each branch. The asterisk after the branch name is used to indicate that the branch is a bottom-level attribute.

In order to calculate the total score for any option on the EFFECTIVENESS node, the analyst must calculate the contribution that each branch makes towards the total score and add them together. (This is true since we assumed independent branches and an additive model.) For example, the score of option I on EFFECTIVENESS would be the score for I on HEAD-UP FLOTATION times the weight of HEAD-UP FLOTATION, plus the score for I on ROUGH SEAS FLOTATION times the weight of ROUGH SEAS FLOTATION, plus the score for I on BUOYANCY times the weight on BUOYANCY, plus the score for I on FAILURE RATE times the weight on FAILURE RATE. Using the numbers in Table 3.4, this would be:

$$\text{SCORE(I)} = (50 \times .35) + (50 \times .10) + (100 \times .38) + (0 \times .17) = 61.$$

Similarly, the score for IB would be:

$$\text{SCORE(IB)} = (100 \times .35) + (100 \times .10) + (100 \times .38) + (70 \times .17) = 95.$$

In general, the score for an option i at a node would be:

$$\text{SCORE(i)} = \sum_j S_{ij} W_j$$

where S_{ij} = score for option i on attribute j;

W_j = weight of attribute j.

The bottom line of Table 3.4, labeled TOTAL, reflects the calculated score for each option on the node EFFECTIVENESS. The total CUMWT reflects the fact that EFFECTIVENESS represents 40% of the entire evaluation model. Note that on EFFECTIVENESS, IB scores highest at 95 while SB scores lowest at 16. Recall that since a relative scoring scale was used, if an option was best on all branches, it would score 100, and, if worst on all, it would score 0.

In a similar fashion, Tables 3.5 and 3.6 show the calculated scores for nodes 1.3 (DURABILITY) and 1.4 (COST). Note that IB scores highest on DURABILITY, but SC scores highest on COST (e.g., cheapest cost).

EXAMPLE FOR ILLUSTRATION ONLY

NODE 1.3, DURABILITY		OPTION SCORES						
BRANCHES:	LOCAL WGT	I	HI	IB	SIB	SB	SC	CUMWT
1.3.1 SHELF LIFE *	.20	0	29	100	100	43	71	.03
1.3.2 EXPOSURE LIFE*	.30	0	17	100	67	33	50	.045
1.3.3 ROUTINE USE *	.40	0	20	100	90	50	60	.06
1.3.4 REUSABILITY *	.10	0	10	100	100	100	100	.015
TOTAL	1.00	0	20	100	86	49	63	.15

Table 3.5: Node 1.3, DURABILITY

EXAMPLE FOR ILLUSTRATION ONLY

NODE 1.4, COST		OPTION SCORES						
BRANCHES:	LOCAL WGT	I	HI	IB	SIB	SB	SC	CUMWT
1.4.1 INITIAL*	.45	38	12	25	0	95	100	.13
1.4.2 O&M *	.55	12	0	100	81	84	94	.17
TOTAL	1.00	24	5	66	45	89	97	.30

Table 3.6: Node 1.4, COST

EXAMPLE FOR ILLUSTRATION ONLY

Before a similar table can be prepared for node 1.2, USABILITY, we recognize that all branches at this node are <u>not</u> bottom-level attributes since COMFORT (1.2.3) is further divided into WEARABILITY (1.2.3.1) and INTERFERENCE (1.2.3.2). The TOTAL line for COMFORT is calculated as shown in Table 3.7:

EXAMPLE FOR ILLUSTRATION ONLY

NODE 1.2.3, COMFORT		OPTION SCORES						
BRANCHES:	LOCAL WGT	I	HI	IB	SIB	SB	SC	CUMWT
1.2.3.1 WEARABILITY *	.60	40	10	0	50	80	100	.045
1.2.3.2 INTERFERENCE*	.40	30	15	0	15	90	100	.03
TOTAL	1.00	36	12	0	36	84	100	.075

Table 3.7: Node 1.2.3, COMFORT

EXAMPLE FOR ILLUSTRATION ONLY

Now, node 1.2 can be addressed as in Table 3.8:

EXAMPLE FOR ILLUSTRATION ONLY

NODE 1.2, USABILITY		OPTION SCORES						
BRANCHES:	LOCAL WGT	I	HI	IB	SIB	SB	SC	CUMWT
1.2.1 IMAGE *	.38	50	20	0	75	90	100	.056
1.2.2 ACCESSIBILITY*	.13	100	80	0	20	90	80	.019
1.2.3 COMFORT	.50	36	12	0	36	84	100	.075
TOTAL	1.00	49	24	0	49	87	98	.15

Table 3.8: Node 1.2, USABILITY

EXAMPLE FOR ILLUSTRATION ONLY

Note that branches 1.2.1, IMAGE and 1.2.2, ACCESSIBILITY are bottom-level as indicated by the asterisk. However, 1.2.3, COMFORT is the aggregate of WEARABILITY and INTERFERENCE, described in Table 3.7. <u>The scores shown for 1.2.3, COMFORT are those calculated in the TOTAL line of Table 3.7</u>. The scores shown for IMAGE and ACCESSIBILITY were directly assessed. The general rules for interpreting scores are as follows:

- For a bottom-level factor, scores are directly assessed;
- For any higher-level factor, scores are calculated using lower-level factors.

We can look at the combined evaluation score considering EFFECTIVENESS, USABILITY, DURABILITY, and COST by examining the matrix shown for node 1.0 PFD as shown in Table 3.9.

Note that the line of scores shown for each branch is directly carried forward from the appropriate TOTAL line in Tables 3.2, 3.3, 3.4, and 3.6, and that the calculation procedure is the same as used at lower-level nodes.

The overall numerical results are found in Table 3.9. The TOTAL line says that IB scores highest (73) followed by SC (66). The reasons why IB did well can be seen easily in Table 3.9.

EXAMPLE FOR ILLUSTRATION ONLY

NODE 1.0, PFD		OPTION SCORES						
BRANCHES:	LOCAL WGT	I	HI	IB	SIB	SB	SC	CUMWT
1.1 EFFECTIVENESS	.40	61	63	95	52	16	33	.40
1.2 USABILITY	.15	49	24	0	49	87	98	.15
1.3 DURABILITY	.15	0	20	100	86	49	63	.15
1.4 COST	.30	24	5	66	45	89	97	.30
TOTAL	1.00	39	33	73	54	53	66	1.00

Table 3.9: Node 1.0, Overall PFD Evaluation

EXAMPLE FOR ILLUSTRATION ONLY

It scored very high (95) on EFFECTIVENESS which was the attribute carrying the most weight. It was highest on DURABILITY (100), and moderate on COST (66). It was worst on USABILITY, but that factor carried a small weight. Similarly, if the analyst wanted to study why IB scored 95 on EFFECTIVENESS, he could look at Table 3.4 to see where points were generated.

(Note: These scores do not represent a Coast Guard evaluation of PFDs. They are hypothetical results used to illustrate MAU methodology.)

The analyst's most important task begins where Table 3.9 leaves off -- he must interpret the results in light of the problem definition and the inputs. In the example problem, the analyst was to identify the major factors in selection of PFDs and evaluate the PFDs for a specified type of decision maker.

After reviewing the numerical results, the analyst can conclude that COST and EFFECTIVENESS are the major differentiating factors with BUOYANCY and HEAD-UP FLOTATION accounting for most of the EFFECTIVENESS differences. If the weights are correct, this result identifies the key areas for further data gathering and sensitivity analysis. All too often, effort is wasted in trying to obtain data on every possible aspect of a problem. It is far more efficient first to identify the critical issues and then gather data only on the factors that can affect the decision. Another important use of the prioritized weights in this case can be to indicate public misperceptions. If the assessed importance weights obtained from a boater survey are at odds with Coast Guard judgment, this could be an indication that boater education programs are in order.

The MAU overall evaluation tells the analyst that for the "profile" of boater described by the weights, the inherent buoyancy PFD would be preferred primarily for its superior effectivess at a moderate cost. The second best option would be seat cushions, primarily due to the boater's desire for a cheaper option. From the Coast Guard's point of view, this can have regulatory policy implications. If the Coast Guard perceives that too many boaters are foregoing effectiveness considerations for lower costs, with the result of more injuries and deaths, tougher laws can be implemented. In this example, the analyst can conclude that inflatables and hybrid inflatables clearly are less preferred alternatives. The analyst should hesitate in making strong inferences about one- or two-point differentials in score; however, a thirty-point differential should safely discriminate among options. The analyst also can use the scores to determine if any options are dominated -- that is, worse than another option on every attribute. In such a case, there is no combination of weights that will cause the dominated option to be preferred. In the PFD example, SR has lower scores than SC on every top-level attribute and would never be preferred to SC under any set of top-level weights.

Before making any firm conclusions, the analyst should test his assumptions through sensitivity analysis. The analyst can study the effects of changing scores, of using a different set of weights to represent a different profile of boater, or of posing a variety of other "what if" questions. Sensitivity analysis will be discussed more fully in Section 3.2.8.

It can't be stressed strongly enough that the numerical answers are not the most important result. These numbers should serve to generate discussion and often, debate. They highlight areas in which results are counterintuitive. Since the judgments behind the resulting numbers are easily traceable, disagreements can focus on specific issues rather than on overall results. When defending the model, if someone disagrees with a number, he should be able to provide strong enough rationale to counterbalance previous judgments. The output of the model is not a decision -- rather, it is a tool to identify principle issues, to focus further data-gathering efforts, and to guide the decisionmaking process.

3.2.8 Sensitivity analysis. Since the judgments behind assessments are often subjective, it is necessary to perform sensitivity analyses on the model inputs. Often, in working with multiple sources of input, there are disagreements that may never be resolved through a consensus building process. Rather than spend significant resources debating the issue, it is better to first determine if a change in the input affects the result. If not, there is little to be gained in further data collection and debate.

There are three major types of sensitivity analyses that are often used. First, the scores that have been assessed can be modified to determine if results change. Experience has shown that results are reasonably insensitive to minor changes in scores and that there is usually a high degree of confidence in the assessed values. Next, several weights can be changed and the overall scores recalculated. This is useful in examining large-scale changes to model (such as using weights for a different decision maker), but does not make it easy to isolate causes of change. A third sensitivity analysis is to vary one weight at a time and identify the regions where decisions change. Typically, one factor is chosen and its weight is allowed to vary from 0% to 100%. As the weight increases, the total weight of the other factors must decrease but the weights are kept in the same relative proportion to each other.

For example, as baseline, the weights at the top level of the PFD model were:

EFFECTIVENESS	40%
USABILITY	15%
DURABILITY	15%
COST	30%

The analyst can examine the effects of letting the weight of COST vary from 0% to 100% of the evaluation. Since the model is linear, the score for any option as a function of COST weight can be plotted on a graph of weight versus overall score. For example, the score for any option can be calculated as follows.

Whatever weight is chosen by COST, the difference between that weight and 100% is reallocated among the other three factors in strict proportion to their original ratio of 40:15:15. Thus, for each weight W_C, assigned to COST, the remaining three weights are calculated as follows:

$$W_E = \frac{40}{40+15+15}(1-W_C) = .57(1-W_C)$$

$$W_U = \frac{15}{40+15+15}(1-W_C) = .21(1-W_C)$$

$$W_D = \frac{15}{40+15+15}(1-W_C) = .21(1-W_C)$$

where the subscripts E, U, and D denote EFFECTIVENESS, USABILITY, and DURABILITY accordingly.

Recalling the formula in Section 3.2.7, the score for alternative I may now be written in terms of W_C as follows:

$SCORE(I)$ = Score for I on EFFECTIVENESS x W_E + Score for I on USABILITY x W_U
+ Score for I on DURABILITY x W_D + Score I on COST x W_C
$= 61 \times .57 \times (1-W_C) + 49 \times .21 \times (1-W_C) + 0 \times .21 \times (1-W_C) + 24 \times W_C$
$= 45.06 - 21.06\ W_C$.

Similarly, scores for other options would be:

$$SCORE\ (HI) = 45.15 - 40.15W_C$$
$$SCORE\ (IB) = 75.15 - 9.15W_C$$
$$SCORE\ (SIB) = 57.99 - 12.99W_C$$
$$SCORE\ (SB) = 37.68 + 51.32W_C$$
$$SCORE\ (SC) = 52.62 + 44.38W_C$$

These can be plotted as a function of W_C as shown in Figure 3.11.

EXAMPLE FOR ILLUSTRATION ONLY

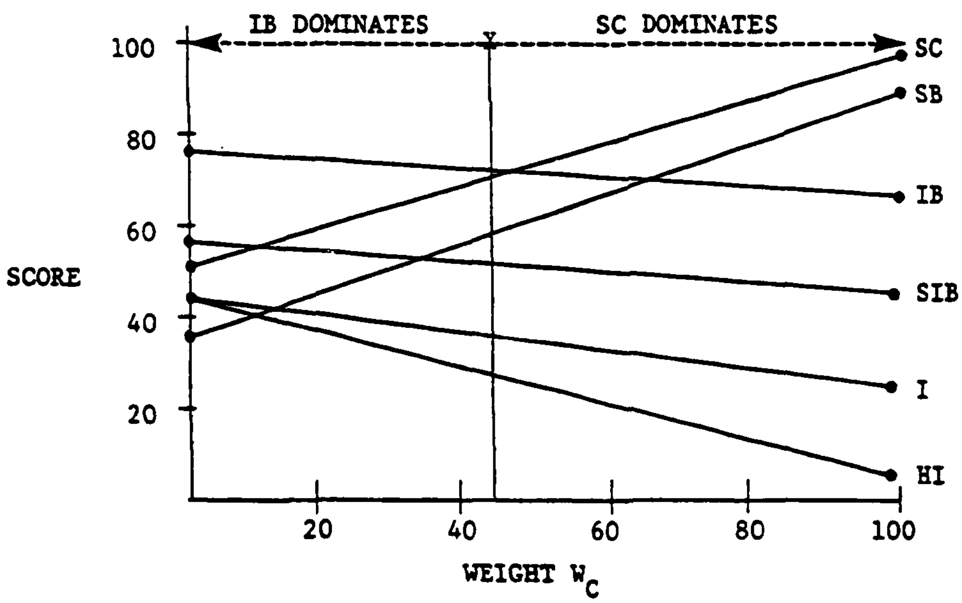


Figure 3.11: Sensitivity on COST

EXAMPLE FOR ILLUSTRATION ONLY

Since we want the option with the highest score, we select the option whose plotted line falls highest. IB is the dominant option if W_C is less than approximately .43, and SC is the dominant option if W_C is greater than .43. We can solve for the exact breakpoint by equating the expressions for the score for these two options:

$$\text{SCORE (IB)} = \text{SCORE (SC)}$$

$$75.15 - 9.15W_C = 52.62 + 44.38W_C$$

$$W_C = .42$$

In the baseline case, $W_C = .30$, and the sensitivity analysis shows that it must be increased to .42 before the preferred alternative changes. Thus a debate as to whether W_C should be .20 versus .30 would not be worth much effort since the result is not affected.

Sensitivity analysis is perhaps the most important step in the MAU process. It helps to solidify subjective judgments and to identify critical areas for further study. It should be an integral part of all MAU analyses.

3.3 Complicating Factors and Extensions

As indicated earlier, the thrust of this chapter is tutorial, and is oriented towards those with minimum experience in applying MAU. While the procedures described in Section 3.2 will be appropriate for the majority of potential applications, there will be occasions when a more complex model is required. This section of the chapter will identify some of these more demanding modeling issues and will provide general references for further reading. Detailed explanation of these complications is beyond the scope of this chapter.

3.3.1 Interdependent evaluation factors. There are occasions when it is not possible to restrict the criteria to be independent of each other without losing a good deal of information. In such cases, the hierarchical structure can include factors that interact, and a multiplicative algorithm can be used to determine an overall score. The analyst should be aware that use of multiplicative models increases the modeling time and effort many fold, and often, the added accuracy provided by such a model is not justified by its costs. A common rule of thumb is known as the 80/20 rule: 80% of the results of an analysis can be achieved with 20% of the input effort. Many experts feel that developing a detailed multiplicative model is tantamount to spending the additional 80% of the input resources to achieve the final 20% of the results. Few decision makers are in a position to afford such luxury.

A detailed description of multiplicative models and their uses can be found in Decisions witn Multiple Objectives: Preferences and Value Tradeoffs by Keeney and Raiffa (see bibliography).

3.3.2 Use of group opinions. The analysis described in preceding sections assumed some source of information. This could be statistical data, analytical models, or judgment. Often, when judgment is the source of information, more than one individual holds an opinion that should be represented in the analysis. In these cases, some thought should be given to determining the best way to elicit and use these opinions. Several ways have been found to be effective in working with group opinions depending on the circumstances.

In cases where it is practical to convene a meeting of appropriate individuals and if these individuals do not hold strongly opposed opinions, then it is practical to try for a group consensus. This might be done by having the analyst lead a group discussion of the problem with the analysis as the focal point. A variation on this method is for the analyst to solicit judgments from a limited number of respondents, develop a complete analysis from these judgments, and then hold a group meeting where this "straw man" analysis is reviewed and refined.

At the other extreme from the group consensus technique are the mechanical methods of combining individual opinions into a single analysis. One such technique is the Delphi method. The Delphi method begins by having individuals give their opinions. Each respondent is then shown all responses but is not told who provided which response. Respondents are then allowed to revise their opinions. After the second round, the responses are averaged or the process is repeated one or more times before averaging. Variations on the Delphi method include simply averaging the first responses or providing feedback on the identities of respondents. These techniques work best when respondents hold different opinions and where it might reasonably be expected that a small group of respondents would inappropriately dominate a group meeting.

A third approach is appropriate when a single decision maker has responsibility for the decision but he wants to be informed of the opinions of others before making the decision. In this case, opinions should be solicited from the appropriate people either individually or in a group meeting. The decision maker should then be informed of these opinions and asked to provide his considered judgments.

Further descriptions of group opinion techniques can be found in Assessment of Group Preferences and Group Uncertainty for Decision Making by David Seaver (see bibliography).

3.3.3 Utility modeling. Utility curves discussed in Section 3.2 were developed using both relative and absolute scoring techniques. There are occasions when it is possible to represent utility judgments in a closed-form equation. One common form is the exponential utility curve described by

$$u(x) = \frac{1-e^{-\gamma}}{1-e^{-\gamma}}$$

where $u(x)$ = utility of x;

γ = risk aversion coefficient (describes risk preferences).

Such utility curves are as shown in Figure 3.12 for varying values of γ.

In cases where a closed-form solution is applicable, the computations required can be less time-consuming and easily adapted to computer algorithms. However, the added complexity makes such models more difficult to communicate to others.

A detailed description of utility modeling can be found in "Risk Preference" by Ronald Howard in Readings in Decision Analysis and in Decision Analysis for the Manager by Brown, Kahr, and Peterson.

EXAMPLE FOR ILLUSTRATION ONLY

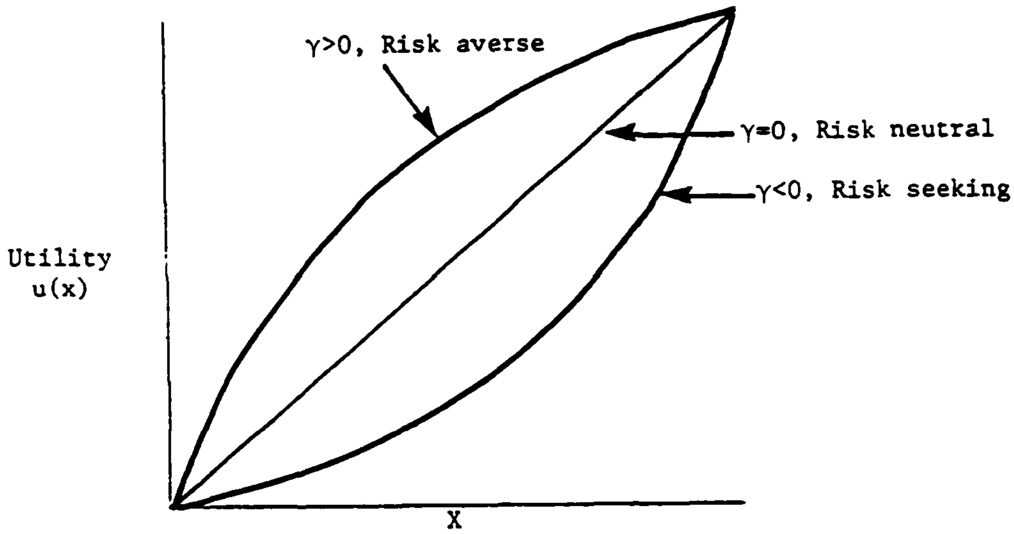


Figure 3.12: Exponential Form Utility Curves

EXAMPLE FOR ILLUSTRATION ONLY

3.3.4 Sensitivity analysis. This manual described techniques for performing sensitivity analysis on one attribute weight at a time. It is possible to allow several weights to vary simultaneously and still plot "decision spaces" as shown in Section 3.2.8. These techniques are time-consuming, but are outstanding as communication tools for interpreting results of the analyses. These techniques are described in Decision Analysis as an Operational Decision Aiding System: Phase III by Peterson, Randall, Shawcross, and Ulvila.

3.3.5 Other applications of MAU In the PFD example, a MAU approach was used for a relatively straightforward evaluation of competing alternatives. There are numerous potential applications of this in a Coast Guard context to include topics such as evaluation of Coast Guard districts, or evaluation of proposed legislation and regulation for boating safety, or evaluation of state safety programs.

There are also applications of MAU in other than an evaluation of alternatives context. One such application is that of a requirements analysis. Typically, current characteristics of an area are examined relative to desired or ideal characteristics, and the "gaps" or deficiencies can be prioritized. Often no alternatives are even discussed. Specific Coast Guard boating safety applications might include requirements analysis for boater education programs, requirements analysis on the need for regulations in unregulated areas, or prioritizing boating scenarios in which accidents occur with an eye towards determining which scenarios require additional attention.

Still another general area for MAU application is that of policy analysis. Rather than evaluating specific regulations, the MAU approach could be used to investigate the impact of policy decisions. Different sets of weights could represent different policy postures, and their effects on boating safety can be examined. One specific application is in the area of policy analysis on minimum safety standards to address issues of who should be the certifying authority and who should be the inspecting authority for re reational boats.

An innovative analyst will find that the MAU approach is a powerful yet flexible tool for analysis, and the range of potential applications can offer considerable assistance to decision makers.

3.4 Summary

MAU modeling techniques can be powerful tools for analyzing complex decisions. They provide a logical framework for decomposing the many interacting aspects of a problem, and allow the analyst to use a "divide and conquer" approach. MAU makes the analysis readily transparent to the decision maker, and provides a convenient audit trail of rationale for all judgments.

Using the techniques described in this chapter, the Coast Guard analysts can address a wide variety of evaluation problems on recreational boating safety. The level of detail in the analysis can be tailored to the level of sophistication of the analyst, but the approach described here is designed for use of those just being introduced to MAU.

While the specific techniques are straightforward, there is no substitute for experience in their application. The analyst should discover the methods that fit most comfortably with his style of analysis, and in time should find himself gravitating towards his own "pet" approaches.

As a final warning, the theory behind MAU is sound and is well documented in the literature; but as with any other technique, if misapplied, the results can be a disaster. The analyst must always remember what the techniques can and cannot do. When properly used, the MAU approach for option evaluation can be highly successful and rewarding.

3.5 Bibliography

Boating statistics 1982. Office of Boating, Public, and Consumer Affairs, U.S. Coast Guard, June 1983.

Brown, R.V., Kahr, A.S., and Peterson, C.R. Decision analysis for the manager. NY: Holt, Rinehart and Winston, 1974.

Cohen, S., Geissler, K., Rossman, H., Ranck, E., and Osciton, R. Recreational boating program effectiveness methodology - A user's manual. Huntsville, AL: Wyle Laboratories, 1981.

Doll, T., Pfauth, et al. Personal flotation devices research - Phase II, (Volume 2). Huntsville, AL: Wyle Laboratories, January 1978.

Handbook for decision analysis. McLean, VA: Decisions and Designs, Inc., 1977.

Keeney, R. and Raiffa, H. Decisions with multiple objectives: Preferences and value tradeoffs. New York, NY: John Wiley and Sons, 1976.

Peterson, C., Randall, L., Shawcross, W., and Ulvila, J. Decision analysis as an element in an operational decision aiding system: Phase III. McLean, VA: Decisions and Designs, Inc., 1976.

Readings in decision analysis. Menlo Park, CA: Stanford Research Institute, 1977.

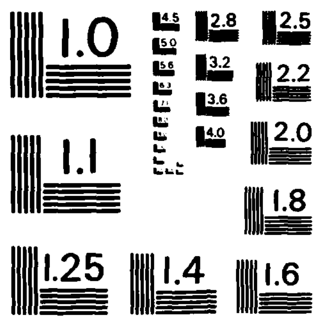
Seaver, D.A. Assessment of group preferences and group uncertainty for decision making. Los Angeles, CA: Social Science Research Institute, 1976.

Ulvila, J.W., and Bresnick, T.A. Resource allocation methods for U.S. Coast Guard Applications (Technical Report 83-7). Falls Church, VA: Decision Science Consortium, Inc., 1983.

Woosley, R.E.D., and Swanson, H.S. Operations research for immediate application. NY: Harper & Row, 1969.

AD-A147 661 RECREATIONAL BOATING SAFETY: ANALYSIS FOR PROGRAMMATIC DECISIONS(U) MANDEX INC MCLEAN VA L GREENBERG ET AL. APR 84 USCG-D-9-84 DTCG23-83-C-20080 2/2

UNCLASSIFIED F/G 13/12 NL



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

4. RESOURCE ALLOCATION METHOD

As noted previously, boating safety analysis embraces four major areas of application. The first three -- Needs Assessment, Performance Prediction, and Performance Evaluation -- formed the framework within which the Wyle report was discussed in Chapter 2.

The fourth major area -- Research Allocation -- is treated in the present chapter. It represents the decision process involved in acting on the results of each (or all three) of the other areas. For example, if Needs Assessment identifies certain areas of boating safety which appear to require corrective action, Resource Allocation identifies which of those areas are the most promising for further exploration, given resource constraints which make it impossible to pursue them all. Similarly, Resource Allocation acts on the results of Performance Prediction and Performance Evaluation to identify which specific mix of proposed and ongoing programs is likely to produce the greatest payoff. In all cases, Resource Allocation seeks to maximize overall utility within defined program constraints.

As in the case of Chapter 3, this chapter explains the technique and outlines the steps involved in applying resource allocation methods to a typical boating safety problem. Again, readers interested in additional detail are referred to the bibliography at the end of the chapter.

4.1 Introduction

Managers often face the problem of how to allocate a scarce resource to different competing uses. The scarce resource might be money or personnel and the uses could be operational or research programs. This report describes a quantitative procedure for allocating resources to provide maximum benefit within a constraint.

The method is described in a straightforward, six-step procedure:

1. Specify options
2. Describe benefit measure

3. Assess benefit across options
4. Assess benefit within options
5. Determine efficient allocations
6. Examine sensitivities.

Each step is described and illustrated in a hypothetical but realistic setting, an allocation of a year's boat testing budget. (All details of the illustration are purely hypothetical. They are intended to illustrate the method, not to solve the problem.)

Although the method is fairly simple, it is powerful enough to be useful in many applied settings where a resource must be allocated. The illustration is the allocation of a financial resource (a year's budget) across testing programs, but the method works as well in other settings. For example, it could be used to help allocate personnel to different Coast Guard task areas. It could even be used to help decide what equipment to add to a boat within a constraint on weight.

The method often appears simple and straightforward in principle but can prove to be complex and elusive in practice. Despite the modest level of mathematics involved, a great deal of skill and expertise may be required to develop a good analysis. These skills cannot be transmitted solely through a report. However, we have included tips which should help a novice analyst to develop an acceptably accurate analysis. These tips are included in the description of the method in Section 4.2 and some special problem areas are discussed in Section 4.3.

The method described in this report is related to the cost-benefit method described by Cohen et al. (1981). They provide a much more detailed presentation of the history and foundations (economic and mathematical) of the method as well as a more detailed and involved discussion of how costs and monetary benefits might be estimated. Our presentation is focused on a simple, practical method that could be used quickly by an analyst. Our presentation is also more general in that we define "benefit" to include any measure that the Coast Guard wishes to maximize. When this method is combined with the multiattribute utility method described in our companion report (Bresnick and Ulvila, 1983), "benefit" can be a multidimensional quantity. (Cohen et al. (1981) use the term "cost-effectiveness" rather than cost-benefit when benefit is non-monetary.) The techniques described in Cohen et al. (1981) are useful supplements to the ones described here in some cases, and the sophisticated analyst may wish to use them.

4.2 Method With Illustration

A cost-benefit method for allocating Coast Guard funds follows six steps, each of which is described in detail in a section below. This description is set in the context of deciding on an allocation of funds to boat testing programs. First, options are specified and costs are estimated. Each program is described in enough detail to assess its probable results and the value of those results. In addition, the specification includes a description of different possible funding levels. Typically, programs could be funded at any of several different levels, and each is described.

Second, a measure of benefit is described. Benefit reflects the goals of the Coast Guard in sponsoring the programs. Benefit might be a single quantity such as improved safety or it may have multiple dimensions such as increased safety, reduced cost to boaters, etc.

Third, benefits are assessed for each program. Each program's relative contribution is assessed using the best data and judgment available.

Fourth, performance is assessed for different levels of funding within each program. For example, some programs degrade in value in exact proportion to funding reductions, others degrade either more rapidly or less rapidly. The rate of degradation of individual programs affects the best mix of funding levels at different levels of total funds.

Fifth, efficient funding allocations are determined. Efficient allocations are generated by identifying the programs (and funding levels within the programs) that offer the greatest increment in benefit per dollar cost. Benefit increments are determined using the information assessed in steps 3 and 4. Cost increments are those assessed in step 1. When plotted, the total benefit from efficient allocations at different levels of total funds will typically look like Figure 4.1.

EXAMPLE FOR ILLUSTRATION ONLY

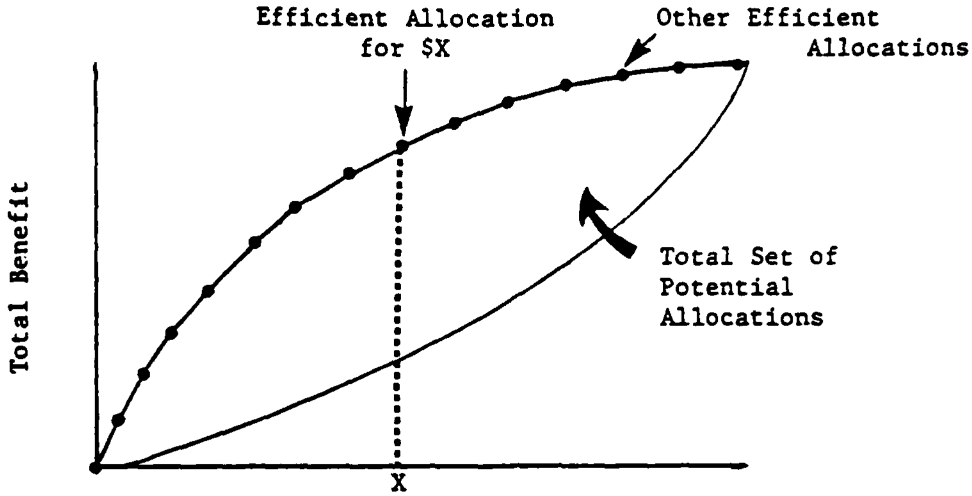


Figure 4.1: Efficient Allocations

EXAMPLE FOR ILLUSTRATION ONLY

Sixth, sensitivity analyses are performed and the model is revised. Efficient allocations can be sensitive to changes in any of the assessments. These might include changes in: cost assessments (step 1), benefit assessments (step 3), and assessments of the relationship between value and funding level (step 4). In addition, some assessments are likely to be highly uncertain or speculative. A sensitivity analysis identifies those parts of the analysis where improvements in the quality of assessment are most important. Another important aspect of the sensitivity analysis is the evaluation of trial packages. Trial packages (proposed sets of funding levels for the programs) are evaluated and compared with efficient allocations. The comparisons indicate the changes required to achieve more value for the same budget (a better allocation).

4.2.1 Describe programs and assess costs. The first step in the method is to describe each program and estimate its cost. Descriptions should be as complete and concise as possible to aid in the assessment of both costs and benefits. These should be written down if possible. In addition, to the extent possible, programs should be described in such a way that each is independent of the rest. This may lead to a combination of otherwise separate programs.

Table 4.1 illustrates descriptions of four testing programs. (This example is used only to illustrate the method, it does not represent Coast Guard policy.) There are four reasonably independent programs:

1. Level flotation and horsepower testing of outboard motor boats;
2. Basic flotation testing of inboard motor boats;
3. Weight capacity testing of outboards;
4. Exploratory testing (subjecting boats to a complete battery of tests).

Notice that the first program, level flotation and horsepower testing, is really a combination of two tests. However, the horsepower test adds so little incremental cost to the level flotation test that it is routinely performed on every boat that is purchased for the level flotation test. For this reason, the programs were combined for this allocation.

Another test, the weight capacity test, also interacts with level flotation. Weight capacity testing is performed on some of the boats purchased for the level flotation test. Costs estimated for the weight capacity test are incremental costs. Thus, the cost of weight testing is increased if insufficient boats have been purchased. In a first cut at the analysis, weight capacity testing is treated as independent, but the dependency is noted. If the specified allocation does not have enough boats, then the analysis must be refined to take the dependency into account. (This problem is discussed in some detail in Section 4.3.3.)

Several possible levels of funding are described for each program. This is especially important as it provides the basis for a pre-analysis of the best changes to make in the event that the budget changes. It also forces a closer scrutiny of the best funding level. It is usually easiest to start with a level that has been proposed or one that is a continuation of previous policy. For example, this might be Level 2 on level flotation. At this level, tests are conducted on 100% of the boats that are highlighted as likely failures (about 80

EXAMPLE FOR ILLUSTRATION ONLY

Table 4.1: Program Descriptions and Funding Levels

Program	Possible Levels of Funding: Minimum	Level 1	Level 2	Level 3
Level flotation and horsepower tests (outboards)	Test 80% of likely failures (64 boats) $320K	Test 100% of likely failures (80 boats) $400K	Test 100% of likely failures, 10 additional boats from major manufacturers and a random sample of 10 additional boats. (100 boats) $500K	Test 100% of likely failures, a boat from every major manufacturer and a representative sample of additional boats (about 100). (200 boats) $1000K
Basic flotation tests (inboards)	Test no boats $0	Test 8 boats $60K	Test 10 boats $75K	Test 20 boats $150K
Weight capacity tests (outboards)	No test $0	Test 25 boats (in addition to level flotation) $18.75K	Test 50 boats (in addition to level flotation) $37.5K	Test 100 boats (in addition to level flotation)* $75K
Exploratory Experiments	None $0	Subject random sample of 40 boats to all tests $240K	Subject random sample of 80 boats to all tests $480K	

*Level 3 on Weight Capacity requires at least Level 2 on Level Flotation.

EXAMPLE FOR ILLUSTRATION ONLY

boats), about 10 additional boats to provide coverage of major manufacturers, and about 10 boats chosen at random, for a total of 100 boats. Level flotation and horsepower tests on 100 outboard motor boats cost the Coast Guard about $500,000, most of which is due to the cost of purchasing the boats.

Additional funding levels are determined by thinking of natural break-points in the number of boats tested. Any of about 4000 new outboard motor boat models might be tested in a given year. Thus, levels could be defined as almost a continuum from 0 to 4000. However, there are some clear principles that can be applied to identify certain points along the continuum for serious consideration. These include the maximum availability of funds, existing criteria for identifying likely failures, and principles of statistical sampling. All three principles contribute to an identification of an upper bound on the test size of about 200 boats. This would include all likely failures, boats from major manufacturers, and a representative sample from remaining boats. This would cost about $1,000,000. Reduced levels of testing would first limit the tests to likely failures, since these 80 boats result in nearly all of the actual failures. An even greater reduction in the program might lead to testing only 80% of the likely failures. Below this level, the entire testing program (including the other three tests) would become suspect, so this is indicated as the minimum level for this test, $320,000.

The levels on other tests were generated by following a similar line of reasoning. In these cases, however, a level of no testing could be included without jeopardizing the value of all other programs, so each had "no test" as its minimum level.

Occasionally, a program will be encountered that does not have natural break-points along the continuum of possible funding levels. For such programs, the analyst should choose a moderate number of levels (between 2 and 8) with cost increments of approximately the size of increments exhibited by other programs.

A comment is also in order concerning cost estimates. The cost estimated should be the cost that is being rationed or allocated. For the illustration, this might be the one-year cost of testing, since the illustration is concerned with the allocation of a year's testing budget. It often helps in identifying the appropriate cost to consider what cost is actually being budgeted. This may be a single year's cost to the Coast Guard or the sum of several years' costs. If several costs are involved, they might be treated as a single cost, for example by discounting. However, to do so assumes that costs can be exchanged, in this case at the discount rate. (For example, if the discount rate is 10% a budget made on discounted costs would imply that an overrun of the year's budget by $10,000 could be made up by being under next year's budget by $11,000.) If this does not hold, this simple method might still provide an adequate approximate result or other more sophisticated methods might be needed (see Section 4.3.2).

4.2.2 Define benefit. The second step in the method is to define the benefit that the Coast Guard expects to obtain from following the set of programs, that which is to be maximized. For some programs, such as the test programs used in the illustration, benefit might be represented by the number of lives saved by the conduct of the programs. Other programs might have other benefits or even multiple benefits. For example, some programs might result in cost savings to Federal, State, or Local governments or to the boating public. Other programs might reduce injuries but not fatalities. Still others might reduce property damage. Other programs might simply contribute to such intangibles as a reduced regulatory burden or to management efficiency. It is important when using this method to be as comprehensive as practical in identifying the benefits of the programs. In cases where benefit is multidimensional, the techniques of multiattribute utility analysis should be used to specify benefit (see Bresnick and Ulvila, 1983). For purposes of this illustration, however, we will assume that lives saved is a sufficient characterization of benefit for the test programs.

4.2.3 Assess benefit for programs. The third step is to assess the benefit contributed by each program. Benefit is assessed in two parts, the relative benefit contributed by different "target level" programs and the change in each program's benefit at different levels of funding. In cases where benefit is a multidimensional quantity, the program's contribution to each dimension must be assessed.

In general, there are a number of possible ways that benefit might be assessed depending on the definition of benefit and on the available data. Some definitions of benefit have natural numerical measures of value such as dollars or lives saved. In these cases, the assessment of benefit is pretty straightforward and assessments are made of the values for the programs. These assessments might need to be modified if value does not vary linearly with the measure (e.g., if the value _per life_ is either greater or less for a large number of lives saved than for a small number), but such adjustments are minor. Another type of benefit definition is one where natural units exist, but the relationship between the units and their value is not obvious. For example, programs might contribute to increasing buoyancy. Buoyancy might be measured in pounds, but the relationship between pounds and value might be non-obvious. In these cases, care must be taken to establish the value of changes in the measure. A third type of benefit is one that has no natural units, for example, projecting a positive image of the Coast Guard. In these cases, benefit might be represented on a relative 100-point scale, where each program's contribution is judged relative to the contribution of other programs.

Introduction of a relative scale is usually useful even when natural scales exist. The relative scale enables easier comparisons across components of benefits if it is multidimensional. In addition, relative scales are often easier to work with if values are assessed judgmentally. The illustration below shows how a natural-unit scale can be transformed into a relative scale.

Benefits might be assessed using data, analysis, judgment, or some combination of these. Data are most useful in estimating the impact of programs on measures that have natural units. For example, statistical data might be used to estimate the number of lives saved by adopting some program. Often, however, appropriate data are not available. For example, statistical data might be available on causes of lost lives, but these data may not indicate how a particular Coast Guard program will save lives. This information can sometimes be provided by analysis, for example using a simulation model. Alternatively, judgment might be used to make the estimate. Judgment is especially important in assessing the impact of programs on relative scales, especially those whose meanings are purely subjective.

Judgment is generally an important element in any assessment of benefit. In the final analysis, value is a subjective concept that can really be estimated only judgmentally. Data might be available to assess the difference in number of lives or dollars saved, but judgment is required to assess how much those savings are worth. Mechanical procedures are often used, such as scaling value proportionally to the units, and these serve as adequate approximations. However, the appropriateness of these procedures is also a matter of judgment.

In the example, illustrative assessments were made of the number of lives that would be saved by conducting each testing program. In this case, estimates might be based on loss-of-life statistics for different types of boating accidents. Estimates are given in Table 4.2. These are estimates for a "perfect" program that would discover all problems of the types tested. Assessments were then needed of the percentage of this benefit that could be expected at each candidate level of funding. These estimates are given in Table 4-3. Notice that this table attributes value in many cases to a program of conducting no tests in the area. Reasoning supporting such assessments might include a deterrent that is provided by the Coast Guard's conduct of any tests (e.g., the minimal level on level flotation and horsepower). Such an assessment might also be highly dependent on whether such omitted tests were conducted in the recent past, and the deterrent value may drop as the time since last conducting the test lengthens. This highlights the importance of the current interpretation of judgments. Judgments that apply to a particular time period (e.g., the next year) may change later.

The assessments in Tables 4.1, 4.2, and 4.3 contain sufficient information to carry out the entire analysis in the example. However, a transformation of this information into a relative benefit scale is useful in other applications, so it will be explained here. The first step in transforming the analysis to a relative value scale is to pick a "target" level for each program, one that will serve as the basis for comparisons within and among programs. For example, target levels for the programs might be as follows:

Program	Target Level
Level flotation and horsepower	Level 2
Basic flotation	Level 2
Weight Capacity	Level 2
Exploratory experiments	Level 1

EXAMPLE FOR ILLUSTRATION ONLY

Table 4.2: Estimated Numbers of Lives Saved by a "Perfect" Program in Each Area

Program	Number of Lives Saved per Year
Level flotation and horsepower	200
Basic flotation	5
Weight Capacity	10
Exploratory Experiments	10

EXAMPLE FOR ILLUSTRATION ONLY

EXAMPLE FOR ILLUSTRATION ONLY

Table 4.3: Percentage of Benefit at Different Levels of Funding

Program	Minimum	Level 1	Level 2	Level 3
Level flotation and horsepower	70%	85%	90%	98%
Basic flotation	70%	80%	90%	98%
Weight capacity	70%	90%	99%	100%
Exploratory experiments	60%	80%	100%	

EXAMPLES FOR ILLUSTRATION ONLY

Table 4.4: Percentages Scaled to Target Level

Program	Minimum	Level 1	Level 2	Level 3
Level flotation and horsepower	0%	75%	100%	140%
Basic flotation	0%	50%	100%	140%
Weight capacity	0%	70%	100%	103%
Exploratory experiments	0%	100%	200%	

EXAMPLE FOR ILLUSTRATION ONLY

Generally, a target level should be the level that is under most serious consideration. Alternatively, the maximum level might be used as the target level when no other level is clearly the most prominent.

Second, the relative benefit is calculated for each program for the improvement between the minimum level and the target level. Using the information in Tables 4.2 and 4.3, these improved benefits are calculated as shown in Table 4.5. For example, the target level of level flotation and horsepower provides 90% of the benefit of a savings of 200 lives/year, and the minimum level provides 70%. Thus, the improvement is 90%-70% = 20% of 200 lives/year or 40 lives/year. To convert these figures to a 100-point relative benefit scale, each figure is divided by 40 (the greatest increment) and multiplied by 100, as shown in the bottom of Table 4.5. (This calculation assumes that value is proportional to number of lives lost.)

Third, the relative benefit is calculated for each level of each program by multiplying the target level's relative benefit by the percentage of the target level achieved, as given in Table 4.4. These calculations are shown in Table 4.6.

It is also of interest to examine the shapes of the curves obtained by plotting the scaled percentage values against funding levels, as shown in Figure 4.2. This figure shows three basically different shapes for three different test programs. The shape of the curve for weight capacity is perhaps the most common. Value rises quickly at lower levels of funding and then levels off. This indicates a decreasing efficiency at higher levels of funding. The shape of the exploratory experiments curve is also common. The straight line indicates a constant level of efficiency. The shape of basic flotation occurs less frequently. This curve rises slowly between $0 and $60,000, quickly between $60,000 and $75,000, and slowly again above $75,000. This indicates that the $60,000 is inefficient, as will be shown by the calculations in Section 4.2.4.

4.2.4 Calculate efficient allocations. The fifth step of the method is to use the assessed benefits and costs to determine efficient allocations, those uses of funds that maximize benefit within a budget. First, lay out a table

EXAMPLE FOR ILLUSTRATION ONLY

Table 4.5: Calculation of Improved Benefits of Target Levels over Minimum Levels

Program	Lives/Year Calculation
Level flotation and horsepower	(.90-.70)(200 lives/year) = 40 lives/year
Basic flotation	(.90-.70)(5 lives/year) = 1.0 life/year
Weight capacity	(.99-.70)(10 lives/year) = 2.9 lives/year
Exploratory experiments	(.80-.60)(10 lives/year) = 2.0 lives/year

Program	Relative Benefit Calculation
Level flotation and horsepower	40 ÷ 40 x 100 = 100
Basic flotation	1.0 ÷ 40 x 100 = 2.5
Weight capacity	2.9 ÷ 40 x 100 = 7.25
Exploratory experiments	2 ÷ 40 x 100 = 5.0

EXAMPLE FOR ILLUSTRATION ONLY

EXAMPLE FOR ILLUSTRATION ONLY

Table 4.6: Relative Benefit at Different Levels of Funding

Program	Minimum	Level 1	Level 2	Level 3
Level flotation and horsepower	0x100=0	.75x100=75	1.0x100=100	1.4x100=140
Basic flotation	0x2.5=0	.50x2.5=1.25	1.0x2.5=2.5	1.4x2.5=3.5
Weight capacity	0x7.25=0	.70x7.25=5.08	1.0x7.25=7.25	1.03x7.25=7.47
Exploratory experiments	0x5=0	1.0x5=5	2.0x5=10	

EXAMPLE FOR ILLUSTRATION ONLY

EXAMPLE FOR ILLUSTRATION ONLY

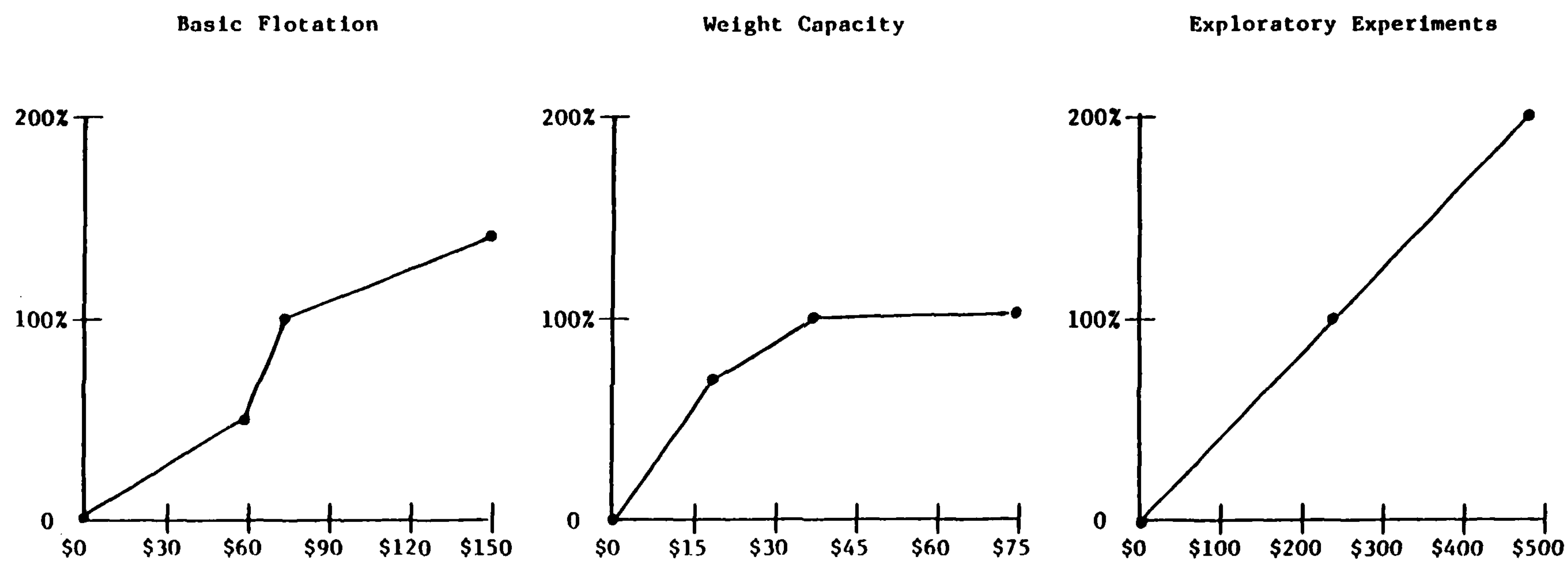


Figure 4.2: Scaled Percentage Value at Different Funding Levels

EXAMPLE FOR ILLUSTRATION ONLY

showing the benefit (from Table 4.6) and cost (from Table 4.1) for each level of each program. Next, calculate the change in benefit and change in cost associated with each change from a lower to a higher level for each program. Then, calculate the ratio of the change in benefit to the change in cost. These operations are shown in Table 4.7. For example, the Level 1 test program for level flotation and horsepower provides 75 more units of relative benefit and costs $80,000 more ($400,000-$320,000) than the minimum level. Thus the incremental benefit to cost ratio is:

$$\frac{\Delta \text{ Benefit}}{\Delta \text{ Cost}} = \frac{75}{\$80,000} = .938 \text{ per thousand dollars.}$$

Other ratios are calculated in a similar manner.

Notice that for the basic flotation program, the benefit-to-cost ratio for moving from Level 1 to Level 2 (.083) is greater than the ratio for moving from the minimum to Level 1 (.021). This corresponds to the dip in the curve in Figure 4.2. For this program, the transition from the minimum to Level 2 is more efficient than the transition from Level 1 to Level 2, so the benefit-to-cost ratio for the transition from the minimum to Level 2 should be calculated and used in the rest of the analysis. This ratio is:

$$\frac{\Delta \text{ Benefit}}{\Delta \text{ Cost}} = \frac{2.5}{\$75,000} = .033.$$

The efficient order for funding the test programs is determined by ordering the transitions between levels on the basis of their benefit-to-cost ratios. The transition with the highest ratio is first, the transition with the second highest ratio is second, and so forth. Choosing to fund the programs in this order ensures maximum benefit within a budget (Everett, 1967). (This statement is approximate if the allocation does not use the entire budget, see Section 4.3.1).

Table 4.8 shows the order of transitions for efficient allocations of funds over the four test programs. For example, the first transition is to increase the size of the level flotation and horsepower test from its minimum of $320,000 to its next level. This gives a benefit of 75 at a total cost of $400,000. The second transition is to increase the level of funding of the weight capacity

EXAMPLE FOR ILLUSTRATION ONLY

Table 4.7: Benefits and Cost for Levels and Transitions

Program		Minimum	Minimum to Level 1	Level 1	Level 1 to Level 2	Level 2	Level 2 to Level 3	Level 3
Level Flotation and Horsepower	Benefit Cost	0 $320K		75 $400K		100 $500K		140 $1000K
	$\frac{\Delta\ \text{Benefit}}{\Delta\ \text{Cost}}$		$\frac{75}{\$80K}=.938$		$\frac{25}{\$100K}=.250$		$\frac{40}{\$500K}=.080$	
Basic Flotation	Benefit Cost	0 $0		1.25 $60K		2.5 $75K		3.5 $150K
	$\frac{\Delta\ \text{Benefit}}{\Delta\ \text{Cost}}$		$\frac{1.25}{\$60K}=.021$		$\frac{1.25}{\$15K}=.083$		$\frac{1.0}{\$75K}=.016$	
Weight Capacity	Benefit Cost	0 $0		5.08 $18.75K		7.25 $37.5K		7.47 $75K
	$\frac{\Delta\ \text{Benefit}}{\Delta\ \text{Cost}}$		$\frac{5.08}{\$18.75K}=.261$		$\frac{2.17}{\$18.75K}=.112$		$\frac{.27}{\$37.5K}=.006$	
Explora-tory Ex-periments	Benefit Cost	0 $0		5.0 $240K		10.0 $480K		
	$\frac{\Delta\ \text{Benefit}}{\Delta\ \text{Cost}}$		$\frac{5.0}{\$240K}=.021$		$\frac{5.0}{\$240K}=.021$			

EXAMPLE FOR ILLUSTRATION ONLY

EXAMPLE FOR ILLUSTRATION ONLY

Table 4.8: Efficient Allocations

Order	Program	Transition From	Transition To	Cost Change*	Benefit Change	$\frac{\Delta B}{\Delta C}$	Total Cost	Total Benefit
0	Level flotation & H.P.		Min				$320	0
1	Level flotation & H.P.	Min	Level 1	$80	75	.938	400	75.0
2	Weight Capacity	Min	Level 1	18.75	4.9	.261	419	79.9
3	Level flotation & H.P.	Level 1	Level 2	100	25	.250	519	104.9
4	Weight Capacity	Level 1	Level 2	18.75	2.1	.112	538	107.0
5	Level flotation & H.P.	Level 2	Level 3	500	40	.080	1038	147.0
6	Basic flotation	Min	Level 2	75	2.5	.033	1112	149.5
7	Exploratory experiments	Min	Level 1	240	5	.021	1352	154.5
8	Exploratory experiments	Level 1	Level 2	240	5	.021	1592	159.5
9	Basic flotation	Level 2	Level 3	75	1.2	.016	1668	160.7
10	Weight capacity	Level 2	Level 3	37.5	.2	.006	1705	160.9

*Note: All costs in thousands

EXAMPLE FOR ILLUSTRATION ONLY

test from $0 to $18,750, which raises the total benefit to 79.9 at a total cost of $418,500. After ten transitions, a total benefit of about 161 is achieved at a cost of $1,705,000. This corresponds to funding each program at its highest level. It is interesting to note that there are only ten transitions in moving from $320,000 to $1,705,000. That is, there are only eleven combinations of funding levels that are efficient (including funding all at minimum levels). This contrasts with the total possible combinations of

$$(4)^3(3) = 192.$$

As the number of programs and levels within those programs increase, the difference between the number of possible combinations and the number of efficient combinations increases dramatically, and the advantages of using this method to select funding levels increases correspondingly. For example, a recent analysis conducted by one of the authors featured twenty-two programs, most of which had three funding levels. The results indicated 65 efficient combinations out of 36,000,000,000,000 possible combinations. For a problem of this size, it is clearly impossible for an analyst to consider all possible combinations. However, the analysis using the method described here was quite manageable.

The results shown in Table 4.8 can be used to specify the best allocation of a budget. This is done by allocating funds to the projects in the order indicated until the budget is exhausted. For example, a budget of $538,000 would fund tests through the fourth order. This allocates $500,000 to level flotation and horsepower testing, $37,500 to weight capacity testing, and nothing to other areas.

It is sometimes useful to examine the results of the analysis by plotting the percentage of total benefit versus total cost for transitions indicated by the efficient order. (Percentage benefit is calculated by dividing the "total benefit" column in Table 4.8 by the maximum total benefit, 160.9.) This plot is shown in Figure 4.3. This shows that value rises most quickly through about the fifth allocation (a little over a million dollars) and then begins to tail off. The curve is also a useful way to display the improvement possible from reallocations of trial packages, as explained in Section 4.2.5.

EXAMPLE FOR ILLUSTRATION ONLY

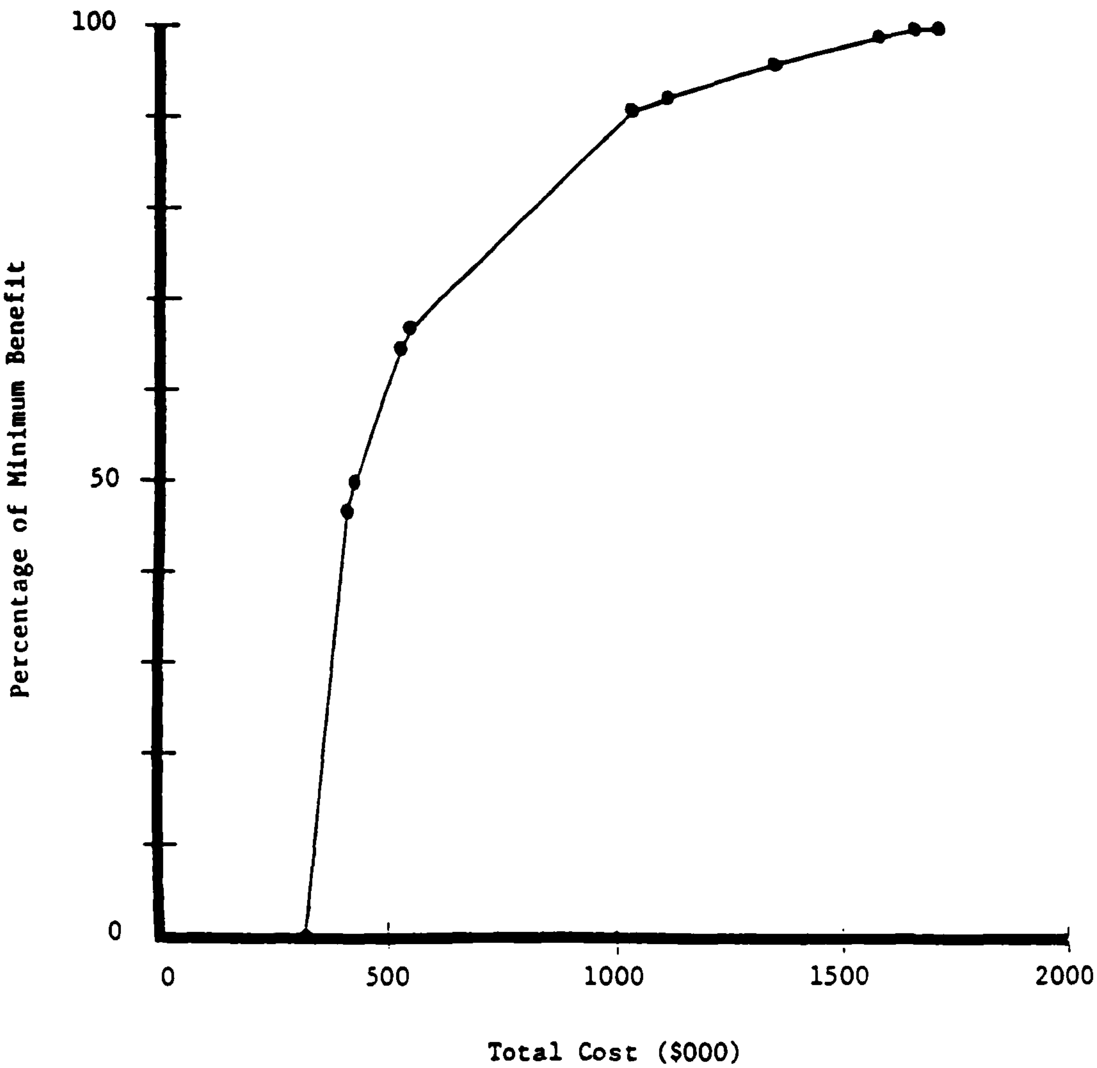


Figure 4.3: Percentage of Maximum Benefit vs. Total Cost for Efficient Allocations

EXAMPLE FOR ILLUSTRATION ONLY

4.2.5 Sensitivity analysis. Often, components of the analysis, cost and benefit assessments, are not known precisely. In addition, some assessments are usually made judgmentally and are subject to disagreement. These conditions make it especially important for an analyst to investigate the sensitivity of the results to variations in the input.

Several inputs might be varied: cost assessments, benefit assessments between programs, or benefit pattern assessments within programs. Each component influences the incremental benefit-to-cost ratio of transitions and thus influences the order of transitions and efficient allocations at different budget levels. Sensitivity analyses might be conducted by varying groups of parameters and re-calculating results or by selectively investigating the extent that certain inputs would need to change in order to give a different result. To change all inputs, the analyst repeats the steps explained above. The discussion below indicates a type of selective variation.

There are two types of changes in the output that the analyst might investigate. One is the change in the entire ordering that results from the change in input. The other is the change in some part of that ordering. The latter investigation is most useful when the analyst has a general idea of the budget level. Here the question becomes, "How sensitive is the budget allocation to changes in input?" Consider the case where the budget is $1,050,000. This is enough to fund through the fifth item. That is, level flotation and horsepower tests are funded at $1,000,000 and weight capacity tests are funded at $37,500. The last increment added to the budget was $500,000 to the level flotation and horsepower test, which has a benefit-to-cost ratio of .080 (from Table 4.7). The next items that would be added with a bigger budget would be a basic flotation test and exploratory experiments, which have benefit-to-cost ratios of .033 and .021.

To test the sensitivity of the result to the inputs for these programs, find the changes needed to change the ordering of their benefit-cost ratios. These calculations are shown in Table 4.9. For example, if there is uncertainty in the benefit assigned to the third level of the level flotation and horsepower test, one could calculate the amount that this would have to be reduced in order

EXAMPLE FOR ILLUSTRATION ONLY

Table 4.9: Sensitivity Analyses

1. Decrease benefit of level flotation and horsepower (Level 3)

A. Compared with basic flotation

$$\frac{B-100}{500} < .033$$

$$B < 116$$

B. Compared with exploratory experiments

$$\frac{B-100}{500} < .021$$

$$B < 110$$

2. Increase benefit to basic flotation or exploratory experiments

A. Basic flotation

$$\frac{B}{75} > .080$$

$$B > 6$$

B. Exploratory experiments

$$\frac{B}{240} > .080$$

$$B > 19.2$$

EXAMPLE FOR ILLUSTRATION ONLY

to make the transition less attractive than basic flotation or exploratory experiments. For basic flotation, the calculation is:

$$\frac{B-100}{500} < .033$$

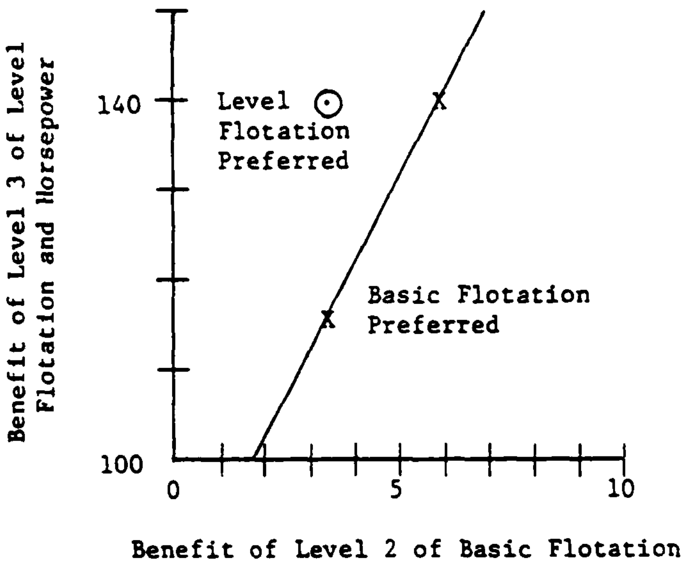
$$B < 116.$$

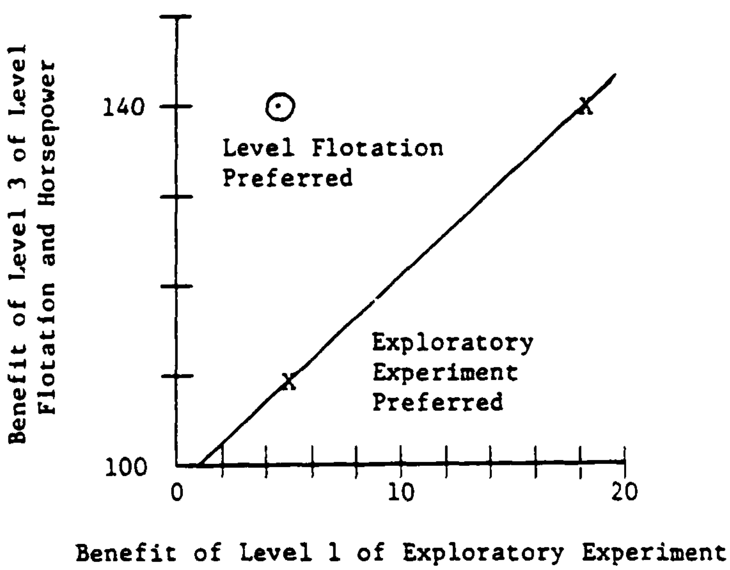
That is, the benefit from the third level of level flotation and horsepower would have to be reduced from its current assessment of 140 to below 116 for basic flotation to be more efficient. If instead, the benefit of basic flotation were varied, its benefit at level 2 would have to increase from its current level of 2.5 to more than 6 to be more efficient. These points can be plotted and connected by a straight line to provide a graphical sensitivity analysis of simultaneous changes in the two inputs as shown in Figure 4.4.

Another type of sensitivity analysis is provided by examining trial packages, or sets of funding levels for all programs. Comparing the benefit and cost of the trial package with that provided by efficient allocations can indicate areas for revision, if the trial package is not efficient. Consider the trial package shown in Table 4.10. This package costs $1,040,000 and provides a benefit of 111.25 (from Table 4.7). An efficient allocation of that budget, however, would fund programs through the fifth transition, providing a benefit of 147 at a cost of $1,037,500. This difference is a substantial fraction of the total possible benefit, as shown in Figure 4.5. This figure also shows that the fourth efficient set provides nearly as much benefit as the trial package, and at a savings of over $500,000. This investigation of the trial package might suggest changes in the analysis. Alternatively, if the trial package is one being considered, the analysis suggests changes to improve efficiency. (Use of trial packages in a sensitivity analysis is especially important with a large number of programs or levels.)

4.2.6 <u>Response to budget changes</u>. The results of the analysis can be used to respond to changes in the budget (or to changes in the constraint, whatever it is). The method determines efficient allocations over the range of budgets, so the response to a budget change is to move down the efficient order if the budget is increased or up if it is reduced. For example, suppose that the

EXAMPLE FOR ILLUSTRATION ONLY





Key: ⊙ is current assessment

Figure 4.4: Graphical Sensitivity Analysis

EXAMPLE FOR ILLUSTRATION ONLY

EXAMPLE FOR ILLUSTRATION ONLY

Table 4.10: Trial Packages and Improvements

Program	Trial Package Funding	Efficient Funding
Level flotation and horsepower	$500K	$1000K
Basic flotation	60K	0
Weight capacity	0	37.5K
Exploratory Experiments	480K	0
Total	$1040K	$1037.5K

EXAMPLE FOR ILLUSTRATION ONLY

EXAMPLE FOR ILLUSTRATION ONLY

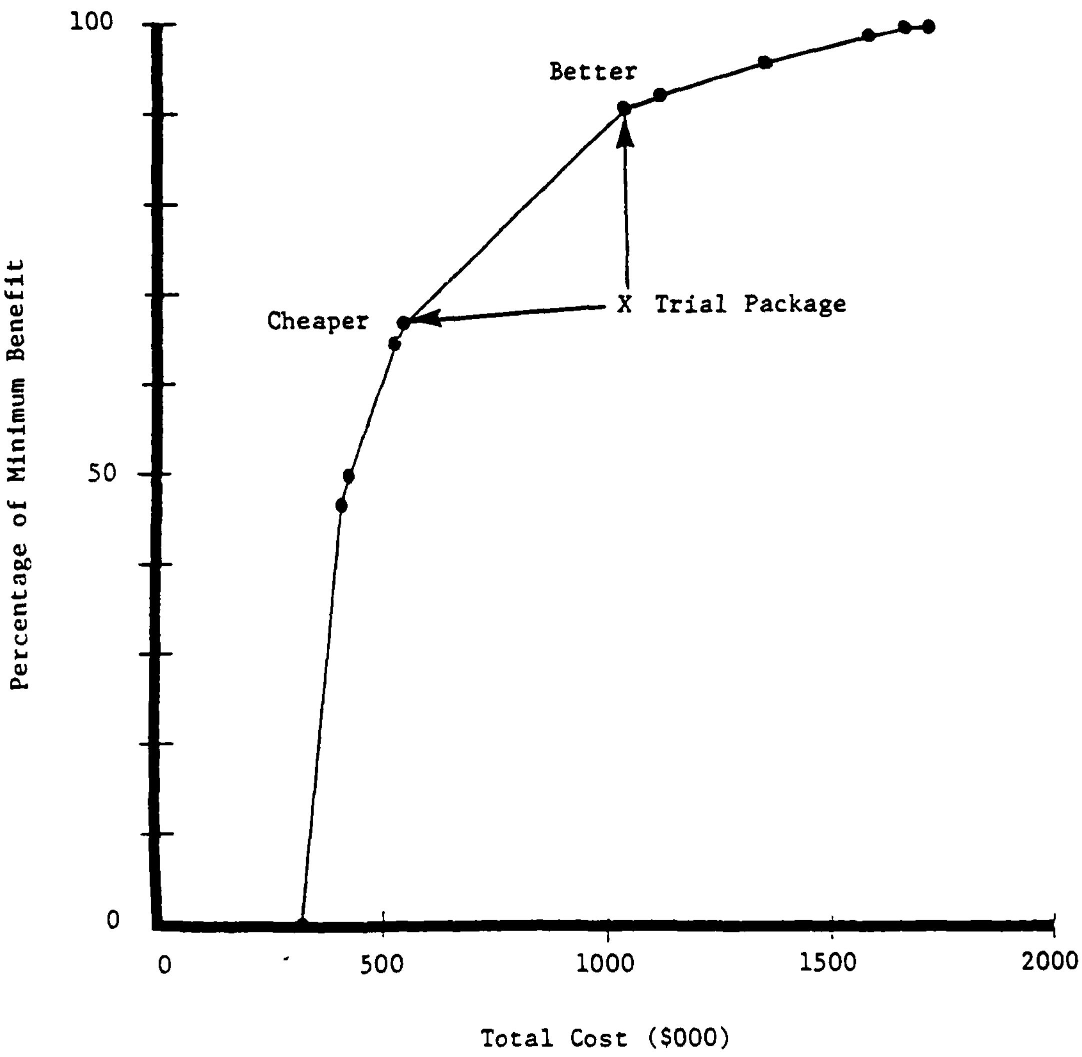


Figure 4.5: Efficient Curve and Trial Package

EXAMPLE FOR ILLUSTRATION ONLY

budget is $1,040,000. Table 4.8 shows that the efficient allocation is to fund programs through the fifth transition. If the budget were increased by $75,000, these funds should go to basic flotation, the sixth in order. If, instead, the budget were increased by $315,000, $75,000 should go to basic flotation and $240,000 to exploratory experiments. If the budget were cut by $500,000, this should come out of level flotation and horsepower. If the budget were increased or decreased to an amount that fell between the identified efficient allocation, then this gives rise to the kind of complication discussed in Section 4.3.1.

4.3 Complicating Factors and Extensions

Section 4.2 describes a method for allocating a single constrained resource so as to achieve maximum benefit. The procedure presented is straightforward and powerful in that it will handle a wide range of problems of this type. However, there are circumstances that complicate the method. Some common ones are discussed here. These include: budgets that lie between efficient sets, multiple constraints, and interdependent programs. In addition, this section presents some extensions of the basic method, uses for other than budgeting, and methods for incorporating a group's judgments into the analysis.

4.3.1 Budgets that lie between efficient allocations. The method provides an ordered list of funding that will maximize benefit within a budget constraint, as shown in Table 4.8. The budget levels ("Total Cost" column) for the allocations, however, are determined by funding levels specified for the programs and may not correspond to the budget. For example, Table 4.8 shows allocations for budgets of $320,000, $400,000, $419,000, $519,000, $538,000, $1,038,000, and so forth. It does not show allocations of $500,000, $750,000, and $1,000,000.

Sometimes, the actual budget may fall close enough to a specified level that the problem is minor. For example, a $500,000 budget may be close enough to the $519,000 that all programs identified in the $519,000 budget can be funded at approximately the specified level, and this may be within the level of estimation error for the costs. This is often the case when a large number of programs are analyzed. For example, Figure 4.6 shows the relative benefit versus cost graph for an analysis with twenty-two programs. Here there were no

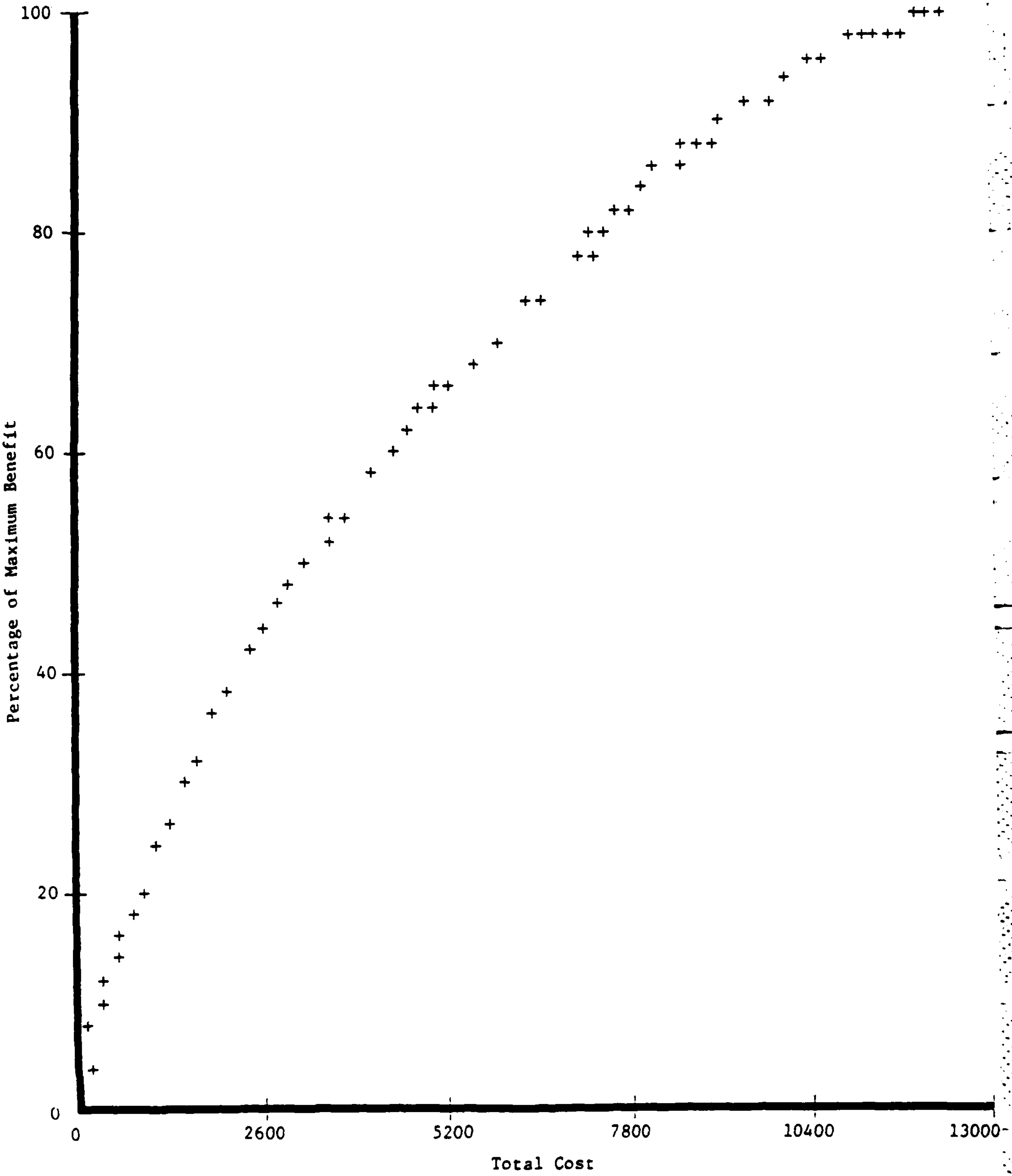


Figure 4.6: Plot of Efficient Allocations with Twenty-Two Programs

large gaps in the total cost of specified efficient funding levels over the range of cost from $0 to $13 million. In this case, an efficient allocation could be found that was close enough to any budget.

At other times, the budget is far from a specified allocation. In our example, a budget of $750,000 is almost half-way between the specified points of $519,000 and $1,038,000. In this case, choosing the funding levels specified at $519,000 leaves $231,000 unused. There are two main ways for dealing with this. First, the analyst should examine the next funding transition that would be made if the budget were available to see if a partial funding would be possible. In this case the next transition would increase funding of the level flotation and horsepower test by $500,000. This would be done by increasing the number of boats tested by 100. Clearly, an intermediate position is possible; 46 additional boats could be tested with an addition of $230,000. If the testing of these additional boats could be expected to provide an additional benefit of 46% or more of that provided by the additional 100 boats, then this should be the allocation. More generally, all programs with large jumps in funding might be reviewed with a view toward splitting up the size of the jumps. The analysis can then be repeated using these newly defined levels of funding.

The second way of dealing with the problem is appropriate when the program level in question is indivisible. For example, suppose that the $500,000 jump was due to the purchase of one piece of that equipment rather than 100 boats. In this case it is not practical to consider buying 46% of the piece of equipment. Here the analyst might fill in the budget by choosing something later in the list that uses most of the remaining funds. Looking at Table 4.8, one sees that the next two transitions are to fund basic flotation at $75,000 and then to fund exploratory experiments at $240,000. Figure 4.3 shows that these two transitions are similar in efficiency. Based on this information, the $231,000 should go to exploratory experiments.

These approximation methods are usually sufficient to overcome this problem. If not, the considerably more complicated technique of integer programming can be used (Bradley, Hax, and Magnanti, 1977).

4.3.2 Multiple constraints. The method described in Section 4.2 is appropriate when a single constraint is imposed, such as budget dollars. Sometimes, the Coast Guard must consider constraints on more than one resource, such as personnel and dollars, when deciding which programs to pursue.

Optimization procedures such as integer programming (Bradley, Hax, and Magnanti, 1977) have long existed for solving such problems. However, these procedures have several drawbacks for this application. First, such a procedure has the appearance of a "black box" that spews forth answers when provided with inputs -- its method of arriving at the answer remains largely a mystery to the user. (Alternatively, the procedure is very complicated to perform manually.) Such a "black box" is likely to be unacceptable given the environment in which the decisions are made. Detailed verbal reasoning is often required to accompany a recommendation, and the "black box" may obscure rather than enlighten such reasoning. In addition, integer optimization techniques may fail to provide a priority list of program funding levels at different budgets. Rather, the optimization indicates the best package of items that can be obtained subject to all constraints, and these packages may change drastically with small changes in the budget level. In particular, there is no guarantee that all of the projects that are recommended with a low budget will continue to be recommended with a higher budget.

A priority list such as that given in Table 4.8, is often a very desirable thing to have. The budget constraint is often subject to a last-minute change; and program funding must respond to the change. The response can be expedited if a priority list is available.

Unfortunately, procedures that optimize subject to multiple constraints do not permit such lists. However, some ad hoc procedures may be useful even though their results are likely to be somewhat sub-optimal. One such procedure works with the cost-benefit list. After the cost constraint is applied, a check is made to see if the other constraints, such as personnel, are exceeded. If they are not exceeded, the analysis remains exactly the same. If, however, the personnel constraint is exceeded, an adjustment might be made in the list by dropping projects with low benefit-to-personnel ratios to the bottom of the list, adding programs with lower benefit-to-cost ratios to the list until both the personnel and cost constraints are met.

Another approach might begin by using a "black box" procedure to determine the best package that meets both the personnel constraint and a tighter-than-necessary budget constraint. Programs in this package form a "high priority" group that is certain to be funded. Additions to this group might then be determined by considering only those remaining items that involve no increase in personnel. These remaining programs could be prioritized according to their benefit-to-cost ratios.

A third procedure involves a redefinition of "cost" to include some mix of personnel and dollars. This procedure then prioritizes the items in accordance with their new benefit-to-"cost" ratios and proceeds much like the standard cost-benefit method. Such a procedure is described in Woosley and Swanson (1969).

All of these ad hoc procedures contain drawbacks that might reduce their usefulness in certain situations. An analyst must use judgment to decide whether the simple method provides a solution that is good enough or whether more complicated methods must be used.

4.3.3 Interdependent programs. The method described in Section 4.2 treats each program as independent in both its cost and benefit. That is, the cost of doing two programs is the sum of their individual costs and the benefit is the sum of individual benefits. This, of course, is often not the case. In fact, the example includes two instances of interdependent programs, each of which was treated in a different way. One instance is the weight capacity test. The costs assessed for this test are incremental costs of testing for weight capacity and do not include the purchase of the boat. It is assumed that a sufficient number of boats will be purchased for the level of flotation test that the cost of testing these boats for weight capacity is just the incremental cost of the weight test. This assumption must be checked. As is seen from the order of funding in Table 4.8, a sufficient number of outboard motor boats are already recommended for the level flotation test each time that weight capacity is indicated, so this assumption holds. If it had been violated, the cost of performing the weight test would have to be increased and the analysis revised. The original analysis also assumes that the criteria used to select boats for the level flotation test are appropriate for the weight test as well. This affects the benefit estimate for the weight test since the value of the weight test is less if performed on inappropriate boats. This assumption was also checked and found to hold.

Another way to handle interdependency is to redefine interdependent programs, combining them into a single program. This was done with the horsepower test and level flotation test. The incremental cost of the horsepower test is very small but its total cost would be large if it involved the purchase of a boat. Thus, the costs of the horsepower and level flotation tests were interdependent. In this case, the two programs were combined into a single one for the analysis.

Either of these two methods might be used to overcome the problem of interdependence. An analyst must use judgment to decide which is best in any given situation. In general, we would recommend the method of combining programs if interdependencies are great and the method of treating the programs as independent if the interdependencies are small.

4.3.4 Extended analysis with other programs. The analysis used in the example included a small number of programs, each with a small number of levels. The analysis can be extended to include many more programs. The most straightforward way to extend the analysis is to simply add the other programs to the list being analyzed and assess their costs and benefits.

Often, programs fall into natural groups characterized by the sponsoring agency or sub-agency or characterized by the type of program. For example, the illustration contained testing programs, but the Coast Guard does its budgeting over other types of programs as well, such as regulatory programs and educational programs. In such cases, it is often easier to compare programs in the same group than to compare programs in different groups. The analysis method can be extended to capitalize on this feature.

The extension involves analyses at two levels. At one level, a separate analysis is performed for all programs within each area. For example, the analysis of test programs might look like the one shown in Section 4.2. At the other level, an analysis is performed across areas. This separation allows the analyst to use different, more appropriate, sources of judgment for the different analysis. For example, personnel responsible for testing might perform the analysis of test programs, those responsible for education might analyze educational programs, those responsible for regulation might analyze regulatory programs. Those people with overall responsibility might then perform the analysis at the higher level, comparing across areas.

The analysis across areas is performed by following essentially the same steps as in the lower level analyses. However, each area is now represented as a single program, and its levels are defined from the output of the lower level analysis. Table 4.11 shows how the testing area might be turned into a program, using the output of the Section 4.2 analysis. First, a choice is made of how many levels to include and which packages to specify as the levels. The figure shows four levels above the minimum. Level 1 corresponds to the funding profile attained by going through the fourth transition on Table 4.8, Level 2 corresponds to the fifth transition, Level 3 to the seventh transition, and Level 4 to the tenth transition. Costs are specified as the total costs. The pattern of benefit is determined by picking a reference level and scaling each level's benefit to the reference level's. In Table 4.11, level 2 is chosen as the reference level. This level has a benefit, as shown in Table 4.8, of 147 and this is scaled to 100%. Using the information in Table 4.8, the relative benefit of the other levels is calculated as follows:

Level 1 = 107 ÷ 147 = 73%
Level 3 = 154.5 ÷ 147 = 105%
Level 4 = 160.9 ÷ 147 = 109%.

In this way, the upper level analysis reflects the lower level evaluations and analysis.

To proceed, similar programs would be developed for the other program areas from the results of their analyses. The analyst would then assess an overall benefit for each program area by comparing reference levels. Results could then

EXAMPLE FOR ILLUSTRATION ONLY

Table 4.11: Program Description Formed from Analysis of Test Programs

Program	Minimum	Level 1	Level 2	Level 3	Level 4
Testing	Level flotation on 64 boats	Level flotation on 100 boats Weight capacity on 50 boats	Level flotation on 200 boats Weight capacity on 50 boats	Level flotation on 200 boats Weight capacity on 50 boats Basic flotation on 10 boats Exploratory experiments on 40 boats	Level flotation on 200 boats Weight capacity on 100 boats Basic flotation on 20 boats Exploratory experiments on 80 boats
Cost	$320K	$538K	$1038K	$1352K	$1705K
Benefit relative to Level 2	0%	73%	100%	105%	109%

EXAMPLE FOR ILLUSTRATION ONLY

be calculated for overall allocations across program areas, which in turn provide the allocations to specific programs. For example, if the overall analysis indicated that testing should be funded at level 3 ($1,352,500), this would indicate that level flotation and weight testing should get $1,000,000, weight capacity should get $37,500, basic flotation should get $75,000, and exploratory testing should get $240,000.

4.3.5 Other uses of the method. The allocation method presented here is illustrated in the problem of allocating a budget to different programs. The method can also be used for other types of problems to determine the best choice within a constraint. One possibility is the allocation of personnel resources instead of dollars. This analysis follows exactly the same steps but personnel are used instead of cost. Another use is to determine the best choice within a physical constraint such as space or weight. For example, it might be used to select the best set of equipment to add to a boat without exceeding some total weight. This analysis follows the same steps, but with weight instead of cost.

4.3.6 Use of group opinions. The analysis described in preceding sections assumed some source of information. This could be statistical data, analytical models, or judgment. Often, when judgment is the source of information, more than one individual holds an opinion that should be represented in the analysis. In these cases, some thought should be given to determining the best way to elicit and use these opinions. Several ways have been found to be effective in working with group opinions depending on the circumstances.

In cases where it is practical to convene a meeting of appropriate individuals and if these individuals do not hold strongly opposed opinions, then it is practical to try for a group consensus. This might be done by having the analyst lead a group discussion of the problem with the analysis as the focal point. A variation on this method is for the analyst to solicit judgments from a limited number of respondents, develop a complete analysis from these judgments, and then hold a group meeting where this "straw man" analysis is reviewed and refined.

At the other extreme from the group consensus technique are the mechanical methods of combining individual opinions into a single analysis. One such technique is the Delphi method. The Delphi method begins by having individuals give their opinions. Each respondent is then shown all responses but is not told who provided which response. Respondents are then allowed to revise their opinions. After the second round, the responses are averaged or the process is repeated one or more times before averaging. Variations on the Delphi method include simply averaging the first responses or providing feedback on the identities of respondents. These techniques work best when respondents hold different opinions and where it might reasonably be expected that a small number of respondents would inappropriately dominate a group meeting.

A third approach is appropriate when a single decision maker has responsibility for the decision but he wants to be informed of the opinions of others before making the decision. In this case, opinions should be solicited from the appropriate people either individually or in a group meeting. The decision maker should then be informed of these opinions and asked to provide his considered judgments. Methods for dealing with group opinions are described in more detail by Seaver, 1976.

4.4 Conclusions

The methods described here can be powerful ones for helping a manager make complex allocation and budgeting decisions and to respond quickly to changing budget conditions. They provide a logical framework for decomposing the aspects of the problem into more manageable pieces and allow the analyst to use a "divide and conquer" approach. These methods make the analysis readily transparent to the decision maker and provide a convenient method for recording and transmitting the rationale for judgments.

Using the techniques described in this chapter, Coast Guard analysts can address a wide variety of resource allocation problems related to recreational boating safety. The level of detail in the analysis and specific analytic methods can be tailored to the level of sophistication of the analyst, but the approach described here is designed to be useful to even a novice analyst.

While the specific techniques are straightforward, there is no substitute for experience in their application. The analyst should discover the methods that fit most comfortably with his style of analysis, and in time should find himself gravitating towards his own "pet" approaches. However, to help introduce an analyst to this variety of techniques, we have included a number of tips and a bibliography.

As a final warning, the theory underlying these techniques is sound and is well documented in the literature; but as with any other technique, if misapplied, the results can be a disaster. The analyst must always remember what the techniques can and cannot do. When properly used, the techniques described here can be highly successful and rewarding.

4.5 Bibliography

Bradley, S.P., Hax, A.C., and Magnanti, T.L. Applied mathematical programming. Reading, MA: Addision-Wesley, 1976.

Bresnick, T.A., and Ulvila, J.W. Multiattribute Utility Analysis for U.S. Coast Guard Applications (Technical Report 83-6). Falls Church, VA: Decision Science Consortium, Inc., 1983.

Cohen, S., Geissler, K, Rossman, H., Ranck, E., and Osciton, R. Recreational boating program effectiveness methodology - A user's manual. Huntsville, AL: Wyle Laboratories, 1981.

Everett III, H. Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. Operations Research, 1967, 399-417.

Seaver, D.A. Assessment of group preferences and group uncertainty for decision making. Los Angeles, CA: Social Science Research Institute of the University of Southern California, 1976.

Woosley, R.E.D., and Swanson, H.S. Operations research for immediate application. NY: Harper & Row, 1969.

END

FILMED

12-84

DTIC